

General Disclaimer

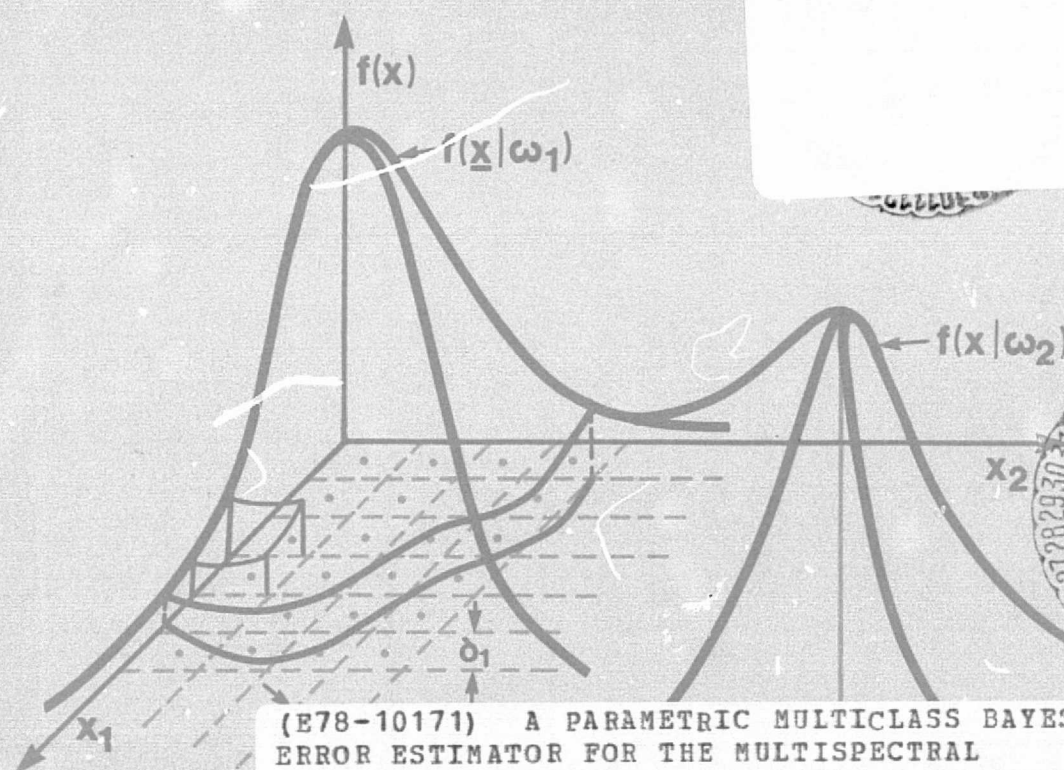
One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
Program information and without liability
for any use made thereof."

A Parametric Multiclass Bayes Error Estimator for the Multispectral Scanner Spatial Model Performance Evaluation

B. G. Mobasseri
C. D. McGillem
P. E. Anuta



(E78-10171) A PARAMETRIC MULTICLASS BAYES
ERROR ESTIMATOR FOR THE MULTISPECTRAL
SCANNER SPATIAL MODEL PERFORMANCE EVALUATION
(Purdue Univ.) 294 p HC A13/MF A01 CSCL 05B

N78-28576

Unclas
00171

G3/43

Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana 47906 USA
TR-EE 78-22 1978

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
Program information and without liability
for any use made thereof."

A PARAMETRIC MULTICLASS BAYES ERROR
ESTIMATOR FOR THE MULTISPECTRAL SCANNER
SPATIAL MODEL PERFORMANCE EVALUATION

B. G. Mobasseri
C. D. McGillem
P. E. Anuta

Published by the
Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana 47906

TR-EE 78-22

This work was sponsored by the National Aeronautics and Space
Administration under Contracts NAS9-14016, 14970 and 15466.

Seay

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	x
GLOSSARY OF SYMBOLS	xvi
ABSTRACT	xx
CHAPTER 1 - Introduction	1
1.1 Statement of the Problem and a Desired Operational Framework	3
CHAPTER 2 - Parametric Bayes Error Estimator in a Multiclass Multidimensional Environment	10
2.1 Review of Previous Work	11
2.1.1 Classification Into Two MVN Distributions With Equal Covariance Matrices	13
2.1.2 Classification Into Two MVN Populations When Covariance Matrices are Unequal	15
2.1.3 Classification Into Multiple MVN Populations	16
2.2 The PMC as a Multiple Integral	18
2.2.1 Decision Boundaries	23
2.3 Approximation to the Classification Error Integral	28
2.3.1 Coordinate Transformation	28
2.3.2 Discrete Space Approach	32
2.4 Error Analysis	37
2.4.1 Statistical Properties of the Estimate	39
2.5 General Comments	48
CHAPTER 3 - Line Scanner Imaging Systems	51
3.1 Types of Systems	52
3.1.1 Multispectral Scanners	52
3.2 System Modeling of a Multispectral Scanner System	54
3.2.1 MSS Spatial Model	56
3.2.2 MSS Statistical Model and Spatial Correlation	59

TABLE OF CONTENTS, cont.

	Page
3.3 Noise in Multispectral Scanner	
System Modeling	85
3.3.1 System Noise	85
CHAPTER 4 - Experimental Evaluation of the Parametric Multiclass Bayes Error Estimator . .	97
4.1 Description of the Data Base	98
4.2 CSP Error Estimation Model:	
Validation and Checkout	99
4.2.1 Fixed Mean, Variable Scatter	104
4.2.2 Fixed Scatter, Variable Mean	160
4.2.3 Classification Error Estimation When the Bayes Rate is Known	176
4.3 Classification Accuracy Estimates Using Landsat and Aircraft MSS Data	184
4.3.1 Ogle County, Illinois	190
4.3.2 Graham County, Kansas	190
4.3.3 Grant County, Kansas	190
4.3.4 Discussion of the Results	193
4.4 Concluding Remarks	194
CHAPTER 5 - Experimental Evaluation of the MSS Spatial Model	197
5.1 Description of the Data Base	198
5.2 Evaluation of the Scanner Characteristic Function	201
5.3 MSS and Classifiability of the Multispectral Data	210
5.3.1 Classification Accuracies at the MSS Output: No Noise	212
5.3.2 Classification Accuracies at the MSS Output: Additive Gaussian Noise	235
5.4 Summary and Conclusions	244
CHAPTER 6 - Conclusions and Observations	247
BIBLIOGRAPHY	251
APPENDICES	
Appendix A: Multispectral Scanner Output Statistics	257
Appendix B: Exponential Spatial Correlation Function Simulator	266
VITA	273

LIST OF TABLES

Table		Page
3-1	Error Matrix for Correlation Function Approximation for Channel 2	66
3-2	Error Matrix for Correlation Function Approximation for Channel 8	67
3-3	Error Matrix for Crosscorrelation Function Approximation Between Channels 2 and 8	68
4-1	Test Cases Arranged by Increasing Separability. Variable Scatter	110
4-2	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 1	111
4-3	Percent CSP and MC Standard Deviations Achieved for Class 1	111
4-4	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 2	115
4-5	Percent CSP and MC Standard Deviations Achieved for Class 2	115
4-6	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 3	119
4-7	Percent CSP and MC Standard Deviations Achieved for Class 1	119
4-8	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 4	123
4-9	Percent CSP and MC Standard Deviations Achieved for Class 1	123
4-10	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 5	127

LIST OF TABLES, cont.

Table		Page
4-11	Percent CSP and MC Standard Deviations Achieved for Class 1	127
4-12	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 6	131
4-13	Percent CSP and MC Standard Deviations Achieved for Class 1	131
4-14	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 7	135
4-15	Percent CSP and MC Standard Deviations Achieved for Class 1	135
4-16	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 8	139
4-17	Percent CSP and MC Standard Deviations Achieved for Class 1	139
4-18	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 9	143
4-19	Percent CSP and MC Standard Deviations Achieved for Class 1	143
4-20	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 10	147
4-21	Percent CSP and MC Standard Deviations Achieved for Class 1	147
4-22	Percent Classification Accuracies Obtained by CSP and MC Estimation Techniques. Case 11	151
4-23	Percent CSP and MC Standard Deviations Achieved for Class 1	151
4-24	Comparison of CSP and MC Percent Classification Accuracy. Variable Scatter	155

LIST OF TABLES, cont.

Table		Page
4-25	Test Cases Arranged by Increasing Separability. Variable Mean	161
4-26	Percent Classification Accuracies Obtained by CSP Estimation Technique. Case 12	162
4-27	Percent Classification Accuracies Obtained by CSP Estimation Technique. Case 13	164
4-28	Percent Classification Accuracies Obtained by CSP Estimation Technique. Case 14	166
4-29	Percent Classification Accuracies Obtained by CSP Estimation Technique. Case 15	168
4-30	Percent Classification Accuracies Obtained by CSP Estimation Technique. Case 16	170
4-31	Percent Classification Accuracies Obtained by CSP Estimation Technique. Case 17	172
4-32	Percent Classification Accuracies Obtained by CSP Estimation Technique. Case 18	174
4-33	Comparison of CSP and MC Percent Classification Accuracy. Variable Mean	177
4-34	Percent Classification Accuracies Obtained by CSP and LARSYS Algorithms, Ogle County, IL	187
4-35	Percent Classification Accuracies Obtained by CSP and LARSYS Algorithms, Graham County, KAN	187
4-36	Percent Classification Accuracies Obtained by CSP and LARSYS Algorithms, Grant County, KAN	188
4-37	Comparison of Percent Classification Accuracies Obtained by CSP and LARSYS Algorithms for Graham County Simulated Data ..	188
4-38	LARSYS Classification Accuracies for Three Realization of Graham County Simulated Data	189

LIST OF TABLES, cont.

Table	Page
5-1 Experimental and Theoretical Scanner Characteristic Function. IFOV=1, Adjacent Line Correlation=0.70	205
5-2 Experimental and Theoretical Scanner Characteristic Function. IFOV=4, Adjacent Line Correlation=0.70	205
5-3 Experimental and Theoretical Scanner Characteristic Function. IFOV=1, Adjacent Line Correlation=0.80	206
5-4 Experimental and Theoretical Scanner Characteristic Function. IFOV=4, Adjacent Line Correlation=0.80	206
5-5 Scanner Output Classification Accuracies vs. IFOV Adjacent Sample Correlation=0.50	214
5-6 Scanner Output Classification Accuracies vs. IFOV Adjacent Sample Correlation=0.55	216
5-7 Scanner Output Classification Accuracies vs. IFOV Adjacent Sample Correlation=0.60	218
5-8 Scanner Output Classification Accuracies vs. IFOV Adjacent Sample Correlation=0.65	220
5-9 Scanner Output Classification Accuracies vs. IFOV Adjacent Sample Correlation=0.70	222
5-10 Scanner Output Classification Accuracies vs. IFOV Adjacent Sample Correlation=0.75	224
5-11 Scanner Output Classification Accuracies vs. IFOV Adjacent Sample Correlation=0.80	226
5-12 Scanner Output Classification Accuracies vs. IFOV Adjacent Sample Correlation=0.85	228
5-13 Scanner Output Classification Accuracies vs. IFOV Adjacent Sample Correlation=0.90	230
5-14 Scanner Output Classification Accuracies vs. IFOV Adjacent Sample Correlation=0.95	232
5-15 Scanner Output Classification Accuracies vs. IFOV SNR=10 dB Adjacent Sample Correlation=0.85	237

LIST OF TABLES, cont.

Table	Page
5-16 Scanner Output Classification Accuracies vs. IFOV SNR=20 dB Adjacent Sample Correlation=0.85	239
5-17 Scanner Output Classification Accuracies vs. IFOV SNR=30 dB Adjacent Sample Correlation=0.85	241

LIST OF FIGURES

Figure		Page
1-1	The Basic Block Diagram of the MSS Model . . .	9
2-1	Allocation of a Measurement Vector X to an Appropriate Partition of the Feature Space	19
2-2	Four Populations with Equal Covariance Matrices: Linear Boundaries	25
2-3	Optimal Partitioning of the Feature Space and the Resulting Quadratic Boundaries .	27
2-4	Three Linear Transformation Steps Prior to Feature Space Sampling	30
2-5	Boundary Location Estimation	33
2-6	A Conceptual Illustration of the CSP Error Estimation Technique Using Two Features . . .	36
2-7	Type II Error Structure	38
2-8	A Case of Net Negative Estimators' Bias . . .	41
2-9	Expected Value of an Area Estimator Using Equidistant Samples. 64 Cells per Axis	43
2-10	A Case of Net Positive Estimator's Bias . . .	44
3-1	MSS Spatial Model as a Linear System	57
3-2	Aircraft MSS Spatial Autocorrelation Function, Channel 2	63
3-3	Aircraft MSS Spatial Autocorrelation Function, Channel 8	64
3-4	Aircraft MSS Spatial Crosscorrelation Function, Channels 2 and 8	65
3-5	Scanner Characteristic Function vs. Scene Correlation Adjacent Line Correlation = .65 . .	74

LIST OF FIGURES, cont.

Figure		Page
3-6	Scanner Characteristic Function vs. Scene Correlation = .7	75
3-7	Scanner Characteristic Function vs. Scene Correlation Adjacent Line Correlation = .75 . .	76
3-8	Scanner Characteristic Function vs. Scene Correlation Adjacent Line Correlation = .8 . .	77
3-9	Scanner Characteristic Function vs. Scene Correlation Adjacent Line Correlation = .85 . .	78
3-10	Scanner Characteristic Function vs. Scene Correlation Adjacent Line Correlation = .9 . .	79
3-11	Scanner Characteristic Function vs. Scene Correlation Adjacent Line Correlation = .95 . .	80
3-12	Scanner Characteristic Function vs. Scene Correlation Adjacent Line Correlation = 1. . .	81
3-13	Scanner Characteristic Function vs. Scene Correlation Adjacent Line Correlation = 0.8 . .	82
3-14	Scanner Characteristic Function vs. Scene Correlation Adjacent Line Correlation = 0.9 . .	83
3-15	Scanner Characteristic Function vs. Scene Correlation Adjacent Line Correlation = 1. . .	84
3-16	Effect of Input Noise on Scanner Output Class Separability	90
4-1	The Block Diagram of the Entire MSS Simulation Procedure	100
4-2	The Configuration of Test Case Mean Vectors Arranged in a Simplex	102
4-3	Small Sample Size Histogram of Simulated Normal Data	108
4-4	Medium Sample Size Histogram of Simulated Normal Data	109
4-5	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 1	112

LIST OF FIGURES, cont.

Figure		Page
4-6	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 1	113
4-7	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 1	114
4-8	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 2	116
4-9	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 2	117
4-10	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 2	118
4-11	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 3	120
4-12	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 3	121
4-13	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 3	122
4-14	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 4	124
4-15	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 4	125
4-16	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 4	126
4-17	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 5	128
4-18	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 5	129
4-19	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 5	130
4-20	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 6	132
4-21	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 6	133

LIST OF FIGURES, cont.

Figure		Page
4-22	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 6	134
4-23	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 7	136
4-24	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 7	137
4-25	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 7	138
4-26	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 8	140
4-27	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 8	141
4-28	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 8	142
4-29	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 9	144
4-30	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 9	145
4-31	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 9	146
4-32	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 10	148
4-33	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 10	149
4-34	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 10	150
4-35	CSP Classification Accuracy Estimate vs. Grid Size. Variable Scatter. Case 11	152
4-36	MC Classification Accuracy Estimate vs. Sample Size. Variable Scatter. Case 11	153
4-37	MC and CSP Error Estimate Standard Deviations. Variable Scatter. Case 11	154

LIST OF FIGURES, cont.

Figure		Page
4-38	CSP Classification Accuracy Estimate vs. Grid Size. Variable Mean. Case 12	163
4-39	CSP Classification Accuracy Estimate vs. Grid Size. Variable Mean. Case 13	165
4-40	CSP Classification Accuracy Estimate vs. Grid Size. Variable Mean. Case 14	167
4-41	CSP Classification Accuracy Estimate vs. Grid Size. Variable Mean. Case 15	169
4-42	CSP Classification Accuracy Estimate vs. Grid Size. Variable Mean. Case 16	171
4-43	CSP Classification Accuracy Estimate vs. Grid Size. Variable Mean. Case 17	173
4-44	CSP Classification Accuracy Estimate vs. Grid Size. Variable Mean. Case 18	175
4-45	Comparison of Bayes and CSP Classification Accuracy Estimates. Two Features	179
4-46	Two Dimensional Partitioned Feature Space . .	182
4-47	One Dimensional Grid Dynamics Illustration . .	183
4-48	Comparison of Bayes and CSP Classification Accuracy Estimates. Four Features	185
4-49	Histogram of Actual Landsat Data. Grant Co., KAN	191
4-50	CSP Classification Accuracy Estimate for Graham Co., KAN	192
5-1	Conceptual Illustration of a Picture Element Viewed from the Satellite	203
5-2	Experimental and Theoretical MSS Characteristic Function Adjacent Line Correlation = .7	207
5-3	Experimental and Theoretical MSS Characteristic Function Adjacent Line Correlation = .8	208
5-4	A Statistical Illustration of the MSS Model .	211

ORIGINAL PAGE IS
OF POOR QUALITY

LIST OF FIGURES, cont.

Figure		Page
5-5	Scanner Output Classification Accuracy vs. IFOV Adjacent Sample Correlation = .5 . . .	215
5-6	Scanner Output Classification Accuracy vs. IFOV Adjacent Sample Correlation = .55 . .	217
5-7	Scanner Output Classification Accuracy vs. IFOV Adjacent Sample Correlation = .6 . .	219
5-8	Scanner Output Classification Accuracy vs. IFOV Adjacent Sample Correlation = .65 . .	221
5-9	Scanner Output Classification Accuracy vs. IFOV Adjacent Sample Correlation = .7 . . .	223
5-10	Scanner Output Classification Accuracy vs. IFOV Adjacent Sample Correlation = .75 . .	225
5-11	Scanner Output Classification Accuracy vs. IFOV Adjacent Sample Correlation = .8 . . .	227
5-12	Scanner Output Classification Accuracy vs. IFOV Adjacent Sample Correlation = .85 . .	229
5-13	Scanner Output Classification Accuracy vs. IFOV Adjacent Sample Correlation = .9 . . .	231
5-14	Scanner Output Classification Accuracy vs. IFOV Adjacent Sample Correlation = .95 . .	233
5-15	Classification Accuracies at the Scanner Output vs. Additive Gaussian Noise and IFOV .	238
5-16	Classification Accuracies at the Scanner Output vs. Additive Gaussian Noise and IFOV .	240
5-17	Classification Accuracies at the Scanner Output vs. Additive Gaussian Noise and IFOV .	242
5-18	Overall Output Classification Accuracy Variation with Noise and IFOV	243

GLOSSARY OF SYMBOLS

a_{ii}	Data spatial autocorrelation model parameter determining the adjacent sample correlation in band i
a_{ij}	Data spatial crosscorrelation model parameter determining the adjacent sample correlation in bands i and j
b_{ii}	Data spatial autocorrelation model parameter determining the adjacent line correlation in band i
b_{ij}	Data spatial crosscorrelation model parameter determining the adjacent line correlation in bands i and j
C_B	Chernoff bound on the probability of correct classification
C	Domain of a sampling cell
c_i	The i th coordinate of a cell's center
CSP	Controlled Space Partitioning
C_{ij}	Cost of making an erroneous allocation
δ_i	Width of a sampling cell along the i th feature axis
Δ^2	Mahalanobis distance
e_i	Error associated with the i th boundary cell
$f(\underline{X})$	Unconditional density function of the measurement vector $\underline{X}$
ϕ_i	Eigenvector matrix for class ω_i
G_S	Grid size in terms of number of cells per axis
Γ_i	Class ω_i domain

Γ_{B_i}	Boundary domain of Γ_i
$h(x,y)$	Two dimensional scanner point spread function
IFOV	Instantaneous field of view
$\bar{J}$	Average Bhattacharyya distance
$\bar{J}^{\%}$	$\bar{J}$ normalized to its highest value
LDF	Linear discriminant function
LR	Likelihood ratio
M	Number of populations in a data set
MC	Monte-Carlo
M_i	Transformed class ω_i mean vector
MVN	Multivariate normal
ML	Maximum Likelihood
μ_i	Mean vector for class ω_i
m_1	Nonzero component of μ_1
N_t	Total number of samples in a data set
N	Dimensionality of the feature space
$N_B^{\%}$	Number of cells around the boundary as a percentage of the cells inside the correct decision domain
N_B	Number of cells along the boundary
n	Number of cells per axis
n_o	Maximum lag in correlation function estimation
n_i	Misclassified samples from class ω_i
ω_i	Class ω_i
PMC	Probability of misclassification
PSF	Point spread function
$\hat{p}_{c \omega_i}$	Conditional estimate of the classification accuracy

P_c	Unconditional classification accuracy
$P(\omega_i)$	Apriori class probability
$P_m(x)$	An orthogonal polynomial
P^2	Total number of pixels in $P \times P$ frame of data
PCM	Pulse code modulation
Q	Complimentary error function
r_{ij}	Spatial correlation coefficient between the i th and j th pixel
R_{pq}	Spatial crosscorrelation matrix for bands p and q
$R_{g_i g_j}$	Scanner's output spatial crosscorrelation matrix for bands i and j
R_g	Entire N band scanner's output spatial correlation matrix
$\rho_{x_k}^i$	Spatial autocorrelation function for cross-track samples in band k
$\rho_{y_k}^i$	Spatial autocorrelation function for along-track samples in band k
ρ_{x_k}	Adjacent sample spatial correlation coefficient
ρ_{y_k}	Adjacent line spatial correlation coefficient
$\rho_{x_{pq}}^i$	Spatial crosscorrelation function for cross-track samples in band p and q
$\rho_{y_{pq}}^i$	Spatial crosscorrelation function for along-track samples in bands
$\rho_{x_{pq}}$	Adjacent sample spatial correlation coefficient in bands p and q
$\rho_{y_{pq}}$	Adjacent line spatial correlation in bands p and q
R^N	N dimensional space of real numbers
r_o	PSF's characteristic length or IFOV

$s_{g_i g_j}$	Spectral correlation coefficient for bands i and j at the scanner's output
$s_{f_i f_j}$	Spectral correlation coefficient for bands i and j at the scanner's input
$\underline{S}_g$	Scanner's output spectral correlation matrix
$\underline{S}_f$	Scanner's input spectral correlation matrix
$\underline{S}_i$	Transformed class ω_i covariance matrix
Σ_i	Class ω_i covariance matrix
v_c	Volume of a sampling cell
W	Anderson's classification statistics
$W_s(\tau, n, a_{ii}, b_{ii})$	Scanner characteristic function for band i
$\hat{W}_i$	Experimentally obtained scanner characteristic function for i th value of adjacent sample correlation
w_i	A set of weighting factors
W_i	Quadratic discriminant functions associated with class ω_i
Z_i	Complement of Γ_i

ABSTRACT

Mobasser, Bijan Gholamreza. Ph.D., Purdue University, May 1978. A Parametric Multiclass Bayes Error Estimator for the Multispectral Scanner Spatial Model Performance Evaluation. Major Professor: C. D. McGillem.

Efficient acquisition and utilization of remotely sensed data requires an extensive apriori evaluation of the performance of the basic data collection unit, the multispectral scanner. The objective is the development of a fully parametric technique to theoretically evaluate the systems response in any desired operational environment and provide the necessary information in selecting a set of optimum parameters.

The probability of correct classification of the various populations in the data is defined as the primary performance index. The multispectral data being of multiclass nature as well, requires a Bayes error estimation procedure that is dependent on a set of class statistics alone. (The underlying problem facing the development of such technique is discussed and a solution based upon sampling of the feature space is proposed.) The classification error estimator is expressed in terms of an N dimensional integral where N is the dimensionality of the feature space. A set of successive linear transformations

prior to the error estimation process provides an N to 1 dimensionality reduction by reducing the Bayes error estimate to a product of N one dimensional integrals. The statistical properties of the estimate is formulated and its relationship with the geometry of the decision boundaries discussed.

The multispectral scanner spatial model is represented by a linear shift-invariant multiple-port system where the N spectral bands comprise the input processes. The scanner characteristic function, the relationship governing the transformation of the input spatial and hence spectral correlation matrices through the systems, is developed. Specific cases for Gaussian and rectangular point spread functions are examined. Random noise is considered and its interpretation in the context of multispectral data is discussed.

In order to validate the Bayes error estimation algorithm's proper performance, multivariate normal data is simulated and the classification accuracy of a set of test cases determined by the parametric and Monte-Carlo type methods. The comparisons of the results provides the required information for evaluation of the theoretical Bayes error estimator performance.

The integration of the scanner spatial model and the parameter classification error estimates provides the necessary technique to evaluate the performance of a

multispectral scanner. A set of test statistics are specified and the corresponding output quantities computed by the characteristic function. Two sets of classification accuracies, one at the input and one at the output is estimated. The scanner's instantaneous field of view is changed and the variation of the output classification performance monitored. The same procedure is followed with additive noise at the scanner output.

In conclusion on the basis of these theoretical results the interaction between the classification accuracy, signal-to-noise ratio, spatial resolution, data spatial correlation and scanner aperture is explained and some suggestions regarding the selection of optimum system parameters is presented.

CHAPTER 1

Introduction

The utilization of earth orbiting platforms as a means of environmental data acquisition has undergone a tremendous growth in the past decade. The feasibility of such techniques was first demonstrated using a multispectral scanner (MSS) carried in a low flying aircraft. The launching of the Earth Resources Technology Satellite (ERTS), later renamed the Landsat, greatly increased the scope of remote sensing technology [1]. Positioned in a polar orbit with a repetitive coverage period of 18 days, a variety of agricultural and environmental data are collected and telemetered to the ground for processing. On board, a rotating mirror multispectral scanner operating in four nonoverlapping bands of electromagnetic radiation constitute the main component of the data collection system [2].

The electromagnetic energy reflected by a target is decomposed into four spectral bands and then transmitted to earth through a PCM channel. The signal degradations caused by various transformations within the scanner subsystem are of great importance. The finite scanner aperture and the atmospheric and quantization noise are but some of the contributing factors. The optimization of the entire set of

interactive parameters within the scanner can be quite involved. From an information processing view, however, five major categories emerge.

1. Spectral band location in the electromagnetic spectrum
2. Spectral bandwidth
3. Number of spectral bands
4. Spatial resolution
5. Signal-to-noise ratio

Due to the finite capabilities of scanner and data analysis techniques, the continuum of the electromagnetic spectrum cannot be fully utilized. Therefore, sampling of the spectrum becomes essential. The band location is generally determined by the target spectral characteristics such that different cover types exhibit different spectral signatures in the same band. The wavelength limits can be shifted somewhat to improve crop identification, but the spectral bandwidth cannot be decreased very much for a fixed signal-to-noise (SNR) ratio. The SNR decreases with decreasing bandwidth.

The spatial resolution has a direct relationship with the signal-to-noise ratio and the classification accuracy. An increase in the resolution requires a narrower aperture which in turn leads to decreased SNR, reduced classification performance and a smaller area scanned for the same data rate. For a coarse resolution, the scanner aperture is wider, SNR is higher classification error rate in general increases, but

'mixed pixels', due to averaging of the adjacent field pixels, will arise.

In a multispectral, remotely sensed data gathering system the final and most important result is the association of each resolution element on the ground with a previously determined population and the evaluation of the performance of such a classification operation. The selection of the parameters within a scanner has as a primary aim the minimization of the probability of misclassification (PMC) of the data. Thus, the classification performance is an indicator against which the choice of other system parameters can be compared.

1.1 Statement of The problem and a Desired Operational Framework

The reflected energy from agricultural and other cover types of interest is corrupted by various noise sources, reshaped by the finite scanner point spread function (PSF) and then quantized and transmitted back to the ground stations for processing. At the Purdue University Laboratory for Applications of Remote Sensing (LARS), the remotely sensed data is analyzed by classifying it into one of M populations by an optimum (minimum probability of error) Bayes classifier. We define the resulting probability of correct classification as the index of performance for a multispectral scanner and describe the goal as the evaluation and simulation of an optimum multispectral scanner system within the framework of

interactive relationships between the spatial resolution, signal-to-noise ratio and classification error rate. Implicit in this statement is the fact that the spectral band location and bandwidth have already been optimally selected as part of the system design process.

The classification accuracy obtained by processing the actual data is necessarily suboptimum due to the aforementioned degradation sources. A reference PMC could be defined by analyzing the performance using the reflected signal at the scanner input, even though this signal is obviously inaccessible. By simulating a theoretical model for the MSS, however, the classification error rate can be evaluated and compared at the scanner input and output thereby establishing an upper bound on the system performance in the context of the defined index of performance. Arbitrary spatial resolution can be specified and its interactive relationship with the SNR and PMC studied.

This interrelationship has been investigated before and some general trends are known. In one experiment [3], initial high resolution aircraft data was classified and the corresponding classification accuracy determined. Lower resolution scanner PSF's were then specified and convolved with the aircraft scanner data to generate a coarser resolution data base. Transformations were carried out for different PSF's and it was concluded that the corresponding PMC's were a decreasing function of the spatial resolution.

The technique employed in [3] is inherently empirical due to the utilization of actual data in the simulation process. Two potential shortcomings of this procedure can be cited: (a), The multispectral signal used is already corrupted by the degradation sources and their effects cannot be isolated; and (b), The accuracy of the system performance calculations is dependent on the size of the available data set. In many applications the data availability can be limited due to the cost, ease of acquisition, availability of ground truth etc. In particular, by convolving an initial data base with a cascade of scanner PSF's to generate a low resolution set, the averaging property of the convolution causes successive reductions in the numerical size of the convolved data and directly affects the statistics and the corresponding estimate of the classification accuracy.

The need for a different algorithm to simulate a multispectral scanner and evaluate its theoretical performance has been demonstrated. This method in order to be as flexible as possible, should depend entirely on the parameters of the model; i.e., population statistics, scanner PSF, noise level etc. Fig. 1-1 is a basic block diagram of the desired MSS model and the performance evaluation process. $\underline{X}$ is the multispectral feature vector. The scanner model is a linear system with specified PSF. The statistical description of the scene is computed at both the scanner input and output, $f(\underline{X}), f'(\underline{X})$. The corresponding probabilities of correct classification are provided by the classification error estimator,

$\hat{P}_C$ and $\hat{P}'_C$. This realization of the process is highly parametric and displays minimum dependence on $\underline{X}$. For a given geographical scene, the scanner PSF and additive noise can be varied and the resulting interaction with $\hat{P}'_C$ observed. Each of the blocks in Fig. 1-1 is composed of various subsystems which will be considered in more detail in later chapters.

The projected algorithm will have several capabilities. The most important one is the ease of parameter manipulation. Variation of the scanner spatial resolution will cause the output statistics to be modified with a corresponding variation in the estimate of the classification error. Similarly, variations in the population separability at the scanner input and the resulting interaction with the PMC can be studied.

This built-in flexibility is a desirable and almost imperative feature of the scanner system modeling. A specific example is the class statistics manipulation. The generation of a new data set, with prescribed statistics, from the existing data set, requires appropriate software and, depending on the data base magnitude, can be potentially time consuming. The alternative in the proposed algorithm is to supply the statistics alone.

The following comment is in order here. Much emphasis has been placed on the data-independent feature of the algorithm. It is clear that this requirement can only be carried so far. Whatever the method, the population statistics must be specified. This condition can be satisfactorily met only by access to an available data set, quality notwithstanding. The distinction emerges at this point that

the contribution of the data to the final result ends at this stage for a parametric model whereas the data utilization will continue throughout the model for an empirical scheme with an error compounding effect.

The MSS simulated model, being a linear system, lends itself to well established system theory methods and, depending on the functions involved, closed form relationships relate the scanner input and output statistics. For the block diagram of Fig. 1-1 to be operational, the contents of the classification error estimator element must be specified. The input to the base is the set of population statistics in the form of M mean vectors and covariance matrices and the output is a set of M performance indices, i.e., the probabilities of correct classification.

The parametric Bayes error estimator is developed in chapter 2. The resulting algorithm requires the data spectral covariance matrices as the only input and produces a set of probabilities of correct classification for each population. In chapter 3 the MSS and multispectral data spatial model is discussed and the desired spectral transfer functions obtained. The experimental results in the form of validation of the classification error estimator using a set of test cases is covered in chapter 4. The scanner spatial model is evaluated in chapter 5 and the associated relationship between the MSS spatial parameters, scene correlation, noise and classification accuracies are discussed.

A summary and suggestions for further work is the topic of chapter 6.

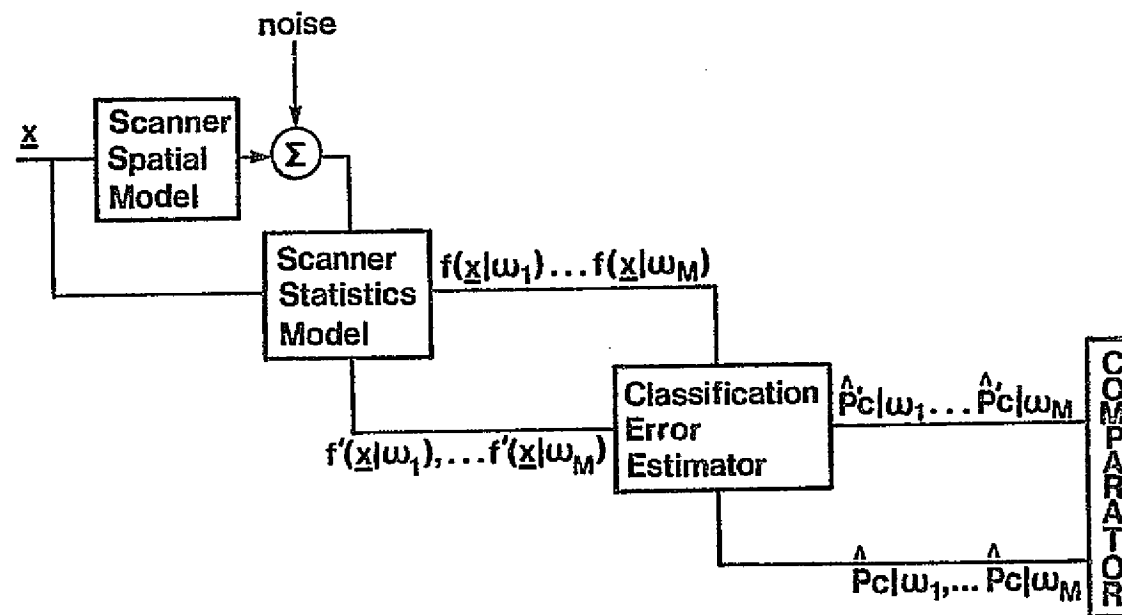


Fig. 1-1 The Basic Block Diagram of the MSS Model.

ORIGINAL PAGE IS
OF POOR QUALITY

CHAPTER 2

Parametric Bayes Error Estimator in a Multiclass Multidimensional Environment

There are basically two types of data classification methods available; parametric; and nonparametric. Nonparametric classification, such as nearest neighbor, is independent of the statistical description of the data, requires access to a large data base and generally is suboptimal relative to the Bayes classifier. It has been shown that the multispectral scanner data can be acceptably described by Gaussian statistics [4]. Therefore, resorting to nonparametric classification would discard valuable a priori knowledge that can improve performance.

Parametric classification, requires the statistical description of the data, either exactly or by parameter estimation. Among all parametric classifiers, Bayes or maximum likelihood (ML) classifiers are optimum in the minimum probability of error sense. Although classification of any data set, parametric or otherwise, is fairly straightforward, determination of the performance of the classifier is far from straightforward. The complexity of the problem is primarily a function of the dimensionality of the measurement space and, to a lesser degree, a function of the multiplicity of populations.

Unless the measurement space is limited to a single dimension, an assumption of very limited applications, exact error rates are not known for ML classifiers.

Multispectral data seldom contains only two classes and always is of a multidimensional nature. The performance calculation for this case has been essentially of a Monte-Carlo nature. The classifier is trained on a portion of the data and then tested on either the same portion or a different segment. The estimate of the probability of error is defined as

$$\hat{\epsilon} = \sum_{i=1}^M P(\omega_i) \frac{n_i}{N_t} \quad (2-1)$$

where M , $P(\omega_i)$, n_i and N_t are the number of populations, a priori class probability, misclassified samples from class ω_i and the total number of available samples, respectively. $\hat{\epsilon}$ is an asymptotically unbiased and consistent estimate of the PMC [5]. Eq. (2-1), with various modifications, is practically the only available PMC estimator. The majority of the literature on statistical classification has been devoted to the case of two multivariate normal populations with heavy emphasis on the equal covariance matrices assumption.

2.1 Review of Previous Work

The field of classification and discrimination, otherwise referred to as allocation, identification, pattern recognition and pattern selection has been one of the most intensely researched areas of statistics and has attracted contributions from a variety of disciplines. In a bibliography,

Anderson et al [6] list over 400 papers published before 1967.

In the beginning stages of research (prior to 1930), the classification problem did not have a precise definition, and was often considered in the context of testing the equality of two distributions [7]. The first clear formulation of the problem is attributed to the pioneering work of Fisher whose ideas were first exposed in the works of other people [8]. In his first paper [9], Fisher considered classification as a problem in multiple taxonomy. For univariate, two population cases he suggested a rule that would assign the measurement $\underline{x}$ to ω_i if $|\underline{x} - \bar{\underline{x}}_i|$ was the smallest of $|\underline{x} - \bar{\underline{x}}_1|$ and $|\underline{x} - \bar{\underline{x}}_2|$ a nearest neighbor rule in current terminology. When measurements were multidimensional, Fisher reduced the problem to the univariate case by selecting a linear combination of the measurements, Fisher's linear discriminant function (LDF), the parameters of which were selected so as to minimize the ratio of the within class scatter to the between class scatter. He called this the optimum linear combination.

One of the most significant developments occurred with the fundamental results of Neyman and Pearson [10]. This was followed by the formulation of the Bayes rule and minimax Bayes rule for two populations and known statistics by Welch [11]. Wald [12] considered the same problem and suggested replacing the unknown quantities with their maximum likelihood estimates. Von Mises [13] obtained a rule that would maximize

the minimum probability of correct classification when an observation is to be assigned to one of several populations.

In a literature survey of a field as diverse as statistical classification, one necessarily has to focus on the particular aspects of the subject most relevant to his work. Therefore, two broad topics; binary group classification and multiple group classification under the assumption of multivariate normal (MVN) statistics are surveyed.

2.1.1 Classification Into Two MVN Distributions With Equal Covariance Matrices

Let the distribution of $\underline{X}$ in ω_i be $N(\underline{\mu}_i, \underline{\Sigma})$ $i=1,2$, where $\underline{\mu}_i$ and $\underline{\Sigma}$ are assumed to be known. This arrangement comprises a classical case for which precise error expressions exist. Let $\Delta^2 = (\underline{\mu}_1 - \underline{\mu}_2)^T \underline{\Sigma}^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$ be the Mahalanobis distance, then

$$P_{\varepsilon|\omega_1} = Q\left(-\frac{\Delta}{2}\right) \quad (2-2)$$

where $Q(\alpha)$ is defined as

$$Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{x^2}{2}} dx \quad (2-3)$$

This case has been discussed by, among others, Welch [11], Wald [12] and Rao [14]. The distribution of classification statistics, if known, can directly provide the error probability. Anderson [15] proposed his W classification statistics by substituting the ML estimates of the unknown parameters in a general likelihood ratio rule (plug-in LR). The

distribution of W proves to be quite complicated to the point of being impractical. Bowker [16] showed that W can be represented as a function of two independent 2×2 Wishart matrices one of which is noncentral. Bowker and Sitgreaves [17] used this result to find the asymptotic expansion of the W distribution function in terms of Hermite polynomials. Teichroew and Sitgreaves [18] used an empirical sampling technique to estimate its distribution. Okamoto [19] considered the statistics of W where the number of degrees of freedom r of $\underline{S}$, the sample pooled covariance matrix, is not necessarily $n_1 + n_2 - 2$, where n_1 and n_2 are the random sample sizes from ω_1 and ω_2 . He then obtained an asymptotic expansion for

$$P[(W - \Delta^2/2)/\Delta < k/\pi_1] \quad (2-4)$$

in terms of n_1 , n_2 and r as n_1 and n_2 tend to ∞ and n_1/n_2 tends to a constant. John [20,21] obtained the distribution of the statistics of W when the common covariance matrix is known.

When the class statistics are based on samples, $T = T(\underline{X}, \underline{\mu}_1, \underline{\mu}_2, \underline{\Sigma})$ is a decision rule whose plug-in estimate, $\hat{T}$, is obtained by substituting the corresponding sample estimates for $\underline{\mu}_1, \underline{\mu}_2$, and $\underline{\Sigma}$, then the conditional error probability based on $\hat{T}$ is given by

$$e_i(\hat{T}) = P[\hat{T} \text{ classifies } \underline{X} \text{ into } \omega_j | \bar{\underline{X}}_1, \bar{\underline{X}}_2, \underline{S}, \underline{\mu} = \underline{\mu}_i] \quad (2-5)$$

$$i, j = 1, 2$$

$$i \neq j$$

The unconditional error probabilities of $\hat{T}$ are $a_i(\hat{T}) = E[e_i(\hat{T})]$. Denote the estimate of $e_i(\hat{T})$ by $\hat{e}_i(\hat{T})$ where the unknown parameters have been replaced by their respective ML estimates. $\hat{a}_i(\hat{T})$ is defined similarly. Let T_0 be the minimax rule with known parameters and $\hat{T}_0$ as its plug-in version. John [22] obtained the distribution of $e_i(\hat{T}_0)$ when $\underline{\lambda}$ is known. Dunn and Varady [23] using an empirical Monte-Carlo technique considered $1 - \hat{a}_i(T_0)$, $1 - e_i(\hat{T}_0)$ and $1 - \hat{e}_i(\hat{T}_0)$ and derived a confidence interval for $e_i(\hat{T}_0)$. Lachenbruch [24] introduced his leaving-one-out method and obtained an almost unbiased estimate for $e_i(\hat{T}_0)$. Hills [25] showed that when $n_1 = n_2$ $a_i(\hat{T}_0) > a_i(T_0)$. Lachenbruch and Mickey [26] compared seven estimation techniques by Monte-Carlo type simulation and concluded that the two most common methods, resubstituting the training samples for testing and the plug-in version of Mahalanobis distance, perform relatively worse than others. Glick [27] showed that as $n_1, n_2 \rightarrow \infty$, $\hat{a}_i(T) \rightarrow a_i(T)$ a.s. uniformly in the class of all rules; moreover, if T is a LR rule, $a_i(\hat{T}) \rightarrow a_i(T)$ a.s. and $\hat{a}_i(\hat{T}) \rightarrow a_i(T)$.

2.1.2 Classification Into Two MVN Populations When Covariance Matrices Are Unequal.

This case differs from the equal covariance matrices case due to the quadratic form of the discriminant function (T being a Bayes rule). Let the distribution of $\underline{X}$ in ω_i be $N(\underline{\mu}_i, \underline{\Sigma}_i)$ $i=1,2$, $\underline{\Sigma}_1 \neq \underline{\Sigma}_2$. Classification statistics again have been a point of interest. Assuming that all the relevant

parameters are known, Cooper [28] studied the optimality of the quadratic discriminant function under stochastic regimes other than normal. When covariance matrices are proportional, Han [29] obtained the distribution of the likelihood ratio and extended the result to circular matrices [30].

Gilbert [31] considered the effect of inequality of the covariance matrices on Fisher's linear discriminant function and concluded that when $\underline{\Sigma}_1 = d\underline{\Sigma}_2$ the performance of Fisher's LDF is adequate only for small values of d . Using simulation techniques, Chaadha and Marcus [32] compared the behavior of three distance statistics and stated that Mahalanobis Δ^2 and Anderson-Bahadur distance are similar in performance and superior to Reyment's generalized distance. Fukunaga and Krile [33], using the distribution of the quadratic discriminant function, expressed the probability of error as an integral and applied the technique to data reported previously [34].

2.1.3 Classification Into Multiple MVN Populations

The problem of optimally classifying an observation into one of M populations under the assumption of general means and covariance matrices and obtaining the error rates has received little attention compared to the previous cases. The reasons are severalfold. Derivation of the classification statistics, so popular in some restricted cases, comes to a halt when faced with the requirement of a joint distribution of M quadratic forms. The solution, if not outright impossible,

is certainly of dubious practical value. Therefore, the assumption of equality of the covariance matrices, accompanied by linearization of the discriminant functions, is widespread. Cacoullos [35] considered the case when the distribution of $\underline{X}$ in ω_i is $N(\underline{\mu}_i, \underline{\Sigma})$ $i=1, \dots, M$ and assigned $\underline{X}$ to the closest ω_i in the Mahalanobis distance sense. Lachenbruch [36] compared the ML rule with Fisher's LDF, the parameters of which are the eigenvalues of a certain matrix. He concluded that when the means are arranged in a simplex, the ML rule performs much better than the LDF and only when the means are collinear is Fisher's LDF performance comparable to the ML method.

In general, multiple group classification is comparatively unexplored, the corresponding error expressions particularly so. In order to make the mathematics tractable, simplifying assumptions have generally been invoked. The assumption of equal covariance matrices reduces the dimensionality of the problem by linearizing the decision boundaries. Hence, an otherwise quadratically partitioned feature space is now divided by hyperplanes. In many cases this can lead to exact error expressions. However, the practicality and usefulness of this procedure is open to question. When the multiple group classification problem is detection of known signals embedded in Gaussian noise, the covariance matrices are indeed equal. In other applications such as classification of various agricultural cover types, however, such an assumption is groundless due to the stochastic nature of the signal itself. The

available error estimation techniques are generally of the empirical Monte-Carlo type.

In addition to the references cited, there are various review articles and bibliographies on classification error estimation. Some of the most comprehensive ones are by Anderson et al [6], Subrahmanian [37], Cacoullos and Styan [38], Lachenbruch [39] and Toussaint [40].

2.2 The PMC as a Multiple Integral

The classification of a multidimensional observation vector into one of M populations is conceptually identical to the binary case. Let Ω , M and N be the feature space, number of classes and the dimensionality of Ω , respectively. The procedure is to divide Ω into M mutually disjoint sets, Γ_i , and to assign each feature vector to a set in accordance with an appropriate rule. This is illustrated in Fig. 2-1.

The estimation of the classification accuracy using the Monte-Carlo technique is possible but frequently undesirable because of accuracy and repeatability limitations and the data dependent nature of the calculation. Therefore, an analytical formulation of the error estimation is sought. Let Z_i , $i=1,2,\dots,M$ partition Ω in R^N . The Bayes risk is defined as [41]

$$R = \sum_{i=1}^M \int_{Z_i} \sum_{j=1}^M P(\omega_j) C_{ij} f(\underline{X}|\omega_j) d\underline{X} \quad (2-6)$$

where C_{ij} is the cost of deciding ω_i where ω_j is true. In

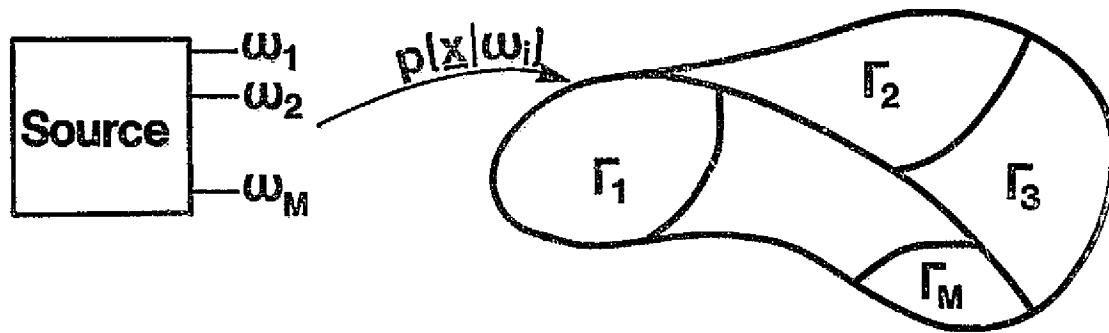


Fig. 2-1 Allocation of a Measurement Vector $\underline{x}$ to an Appropriate Partition of the Feature Space.

the case where $C_{ij} = 0$ for $i=j$ and $C_{ij} = 1$ for $i \neq j$ R is the probability of error.

Among all of the possible choices of Z_i , the Baye rule partitions Ω into $Z = Z^*$ such that $R = R^*$ is the minimum probability of error [41]. Assuming that the population statistics follow multivariate normal law, the optimum Bayes rule is as follows [42]

$$\underline{X} \in \omega_i \quad \text{if } W_i < W_j \quad \forall_i \neq j \quad i=1,2,\dots,M$$

where

$$W_i = (\underline{X} - \underline{\mu}_i)^T \underline{\Sigma}_i^{-1} (\underline{X} - \underline{\mu}_i) + \ln |\underline{\Sigma}_i| - 2 \ln P(\omega_i)$$

with $\underline{X}$ = observation vector

$\underline{\mu}_i$ = mean vector for class ω_i

$\underline{\Sigma}_i$ = covariance matrix for class ω_i

$P(\omega_i)$ = a priori probability for class ω_i

(2-7)

The error estimate based on direct evaluation of (2-6) exhibits all the desired properties outlined previously.

The evaluation of multiple integrals bears little resemblance to their one dimensional counterparts, mainly due to the vastly different domains of integration. Whereas there are three distinct regions in one dimension; finite, singly infinite, and doubly infinite; in an N dimensional space there can be potentially an infinite variation of domains. Thus, the established one dimensional integration techniques do not, in general, carry over to an N dimensional space. Therefore, it is not surprising that no systematic technique exists for the evaluation of multivariate integrals. The available

methods are generally applicable to elementary regions and integrals.

Let us examine the domains of integration encountered in the Bayes error estimation. The regions of integration, Γ_i , are defined by the inequality $W_i < W_j \quad \forall j \neq i$. Therefore, Γ_i is defined by a set of intersecting hyperquadratics, the mathematical representation of which is too complicated to be practical. The population statistics, of course, determine the geometrical shape of a boundary. The most tractable geometry results from the assumption of identical covariance matrices, $\underline{\Sigma}_i = \underline{\Sigma} \quad \forall i$. An orthonormal transformation reduces $\underline{\Sigma}$ to an identity matrix; hence, each discriminant function W defines a hypersphere centered at the population mean in the transformed coordinate system. Such an arrangement leads to hyperplanes as optimum partitions of the feature space.

The assumption of equal covariance matrices, albeit unrealistic, is prevalent in the statistical classification literature and has its roots in the linear property of the boundaries. Fig. 2-2 shows a case of four populations with two features.

In approximating the solution to any multiple integral, the parameters to be determined are a set of weighting factors, $w_1, w_2, \dots, w_m$ and a set of points, $p_1, p_2, \dots, p_n$ in Z . Then (2-6) can be represented as a finite Riemann sum

$$\int_Z f(\underline{X}) dV = \sum_{i=1}^m w_i f(p_i) \quad (2-8)$$

In order to illustrate the difficulties involved with evaluating (2-8), an examination of its one dimensional counterpart

$$\int_a^b k(x)f(x)dx \approx \sum_{i=1}^m w_i f(p_i) \quad (2-9)$$

where $k(x)$ is a weighting function, is useful. One way to evaluate (2-9) is to pre-select p_i according to a certain rule and require that $w_1 \dots w_m$ be chosen such that

$$e = \int_a^b k(x)f(x)dx - \sum_{i=1}^m w_i f(p_i) \quad (2-10)$$

is zero for all monomials of degree n . The Newton-Cotes integration technique is a prime example of this rule where the interval (a,b) is divided into m equal subintervals of length $(b-a)/m$. Among other well known methods having this property are the trapezoidal and Simpson's rule.

Sometimes, it is advantageous to have a set of points with unequal spacing. The most common choice is when $p_1, \dots, p_m$ are the m zeros of an orthogonal polynomial $P_m(x)$. There are numerous methods each using a particular set of polynomials to generate the desired abscissas [43], among them are the Chebyshev orthogonal polynomials of the first and second kind. This approach provides a relation similar to (2-9) except that the rule is exact for all polynomials of degree $2m-1$. A notable example is the m -point integration rule of the Gauss Type.

The extension of one dimensional techniques to higher dimensional spaces is hindered for a variety of reasons. As pointed out previously, orthogonal polynomials play an

important part in the evaluation of one dimensional integrals; however, there is no generalization of such method to higher dimensions. For example, given m points $p_1, \dots, p_m$ in R^2 it may not be possible to find a polynomial, in x and y , to take on prescribed values at the points p_i . The next item is the more complicated structure of the N -dimensional functions which necessarily causes complicated domains of integration such as (2-7).

2.2.1 Decision Boundaries

In a series of papers, Cooper explored various decision boundaries arising in a pattern classification problem, with the emphasis on the optimality of some well known rules under more general conditions. Hyperspheres arising from spherically symmetric distributions were found to be optimum for Pearson Type II and Type VII in addition to a normal distribution [44,45]. Error expressions were obtained by integration of the random measurement vector $|X|$ within the constant radius sphere. For the more general case, quadratic partitions were claimed to be optimum for not only normal population but for the general class of monotone distributions with equal determinant covariance matrices [46]. In the latter case, the statistics, not the functional form of the class density functions, are the only required parameters.

Although multiclass, multifeature data classification is straightforward, the probability of error estimation through non-Monte Carlo techniques shows only a structural

similarity to the binary, unidimensional case. It is significant that the complexity of the classification accuracy estimation in the general case under study is mainly a function of the dimensionality of the feature space and only partly a function of the population multiplicity. An exact error expression, for example, exists for M-class, single dimension Bayesian classification.

The integral expressing the error is an Nth order multiple integral over domains defined by (2-7). Consider Fig. 2-2. The region of interest which would yield the highest probability of correct classification for class ω_1 is a triangle.

Let

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} 0 \\ m \end{bmatrix} \quad \mu_3 = \begin{bmatrix} -m \\ -m \end{bmatrix} \quad \mu_4 = \begin{bmatrix} m \\ -m \end{bmatrix}$$

then Γ_1 is a set defined by the following simultaneous inequalities

$$\Gamma_1: \begin{cases} x_2 < m/2 \\ x_2 > -(x_1 + m) \\ x_2 > x_1 - m \end{cases}$$

Hence

$$\hat{P}_{c|\omega_1} = \int_{\Gamma_1} f(\underline{X}|\omega_1) d\underline{X}_1 = 2 \int_0^{\frac{3m}{2}} \int_{x_1-m}^{\frac{m}{2}} f(x_1, x_2|\omega) dx_1 dx_2 \quad (2-11)$$

Therefore, $\hat{P}_{c|\omega_1}$ can be evaluated to any degree of precision desired. The point of this simple example was to demonstrate the importance of the boundaries of Γ_i . The ease of formulation

ORIGINAL PAGE IS
OF POOR QUALITY

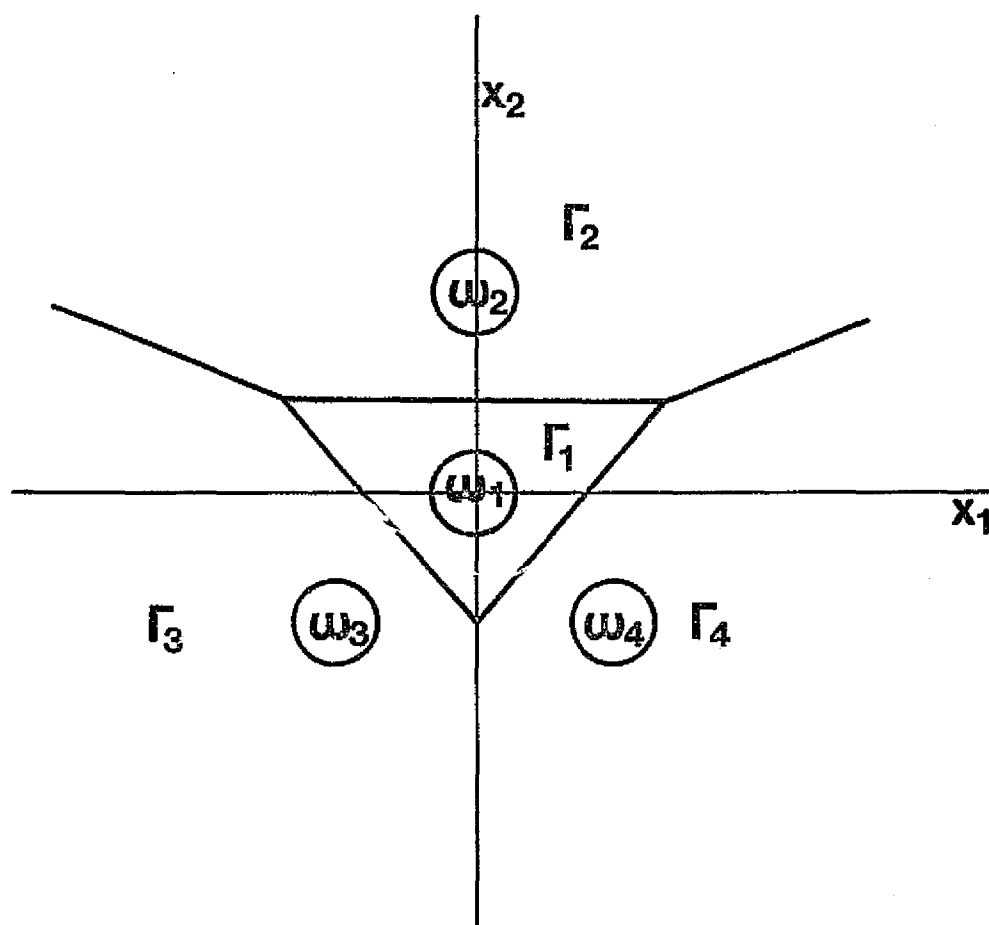


Fig. 2-2 Four Populations with Equal Covariance Matrices: Linear Boundaries.

was mainly due to the linear contours of integration, precipitated by the equal covariance matrices.

Relaxing the equal covariance matrices assumption considerably complicates the problem. First, there is no transformation, unitary or otherwise, that would decouple the feature space for all the populations simultaneously; and second, the boundaries of interest are now portions of various hyperquadrics. These two changes alone would rule out any meaningful representation of $\hat{P}_C$ in a form similar to (2-11). Fig. 2-3 shows a typical multiclass case.

The dimensionality of feature space can be regarded as the most important complicating factor. There are at least three parameters dependent on N .

1. order of the error integral
2. geometry of Γ_i 's
3. computation time

The existence (or lack of it) of techniques in evaluation of multiple integrals has been discussed before. While it could be argued that the multiplicity of populations, M , has a more pronounced affect on the decision regions, it is undoubtedly true that there are *no* complex boundaries in one dimension regardless of the value of M . In addition, boundary visualization, so helpful in error estimation, will no longer be possible for $N \geq 3$. It will be shown in later chapters that the computation time is related exponentially to N and linearly to M .

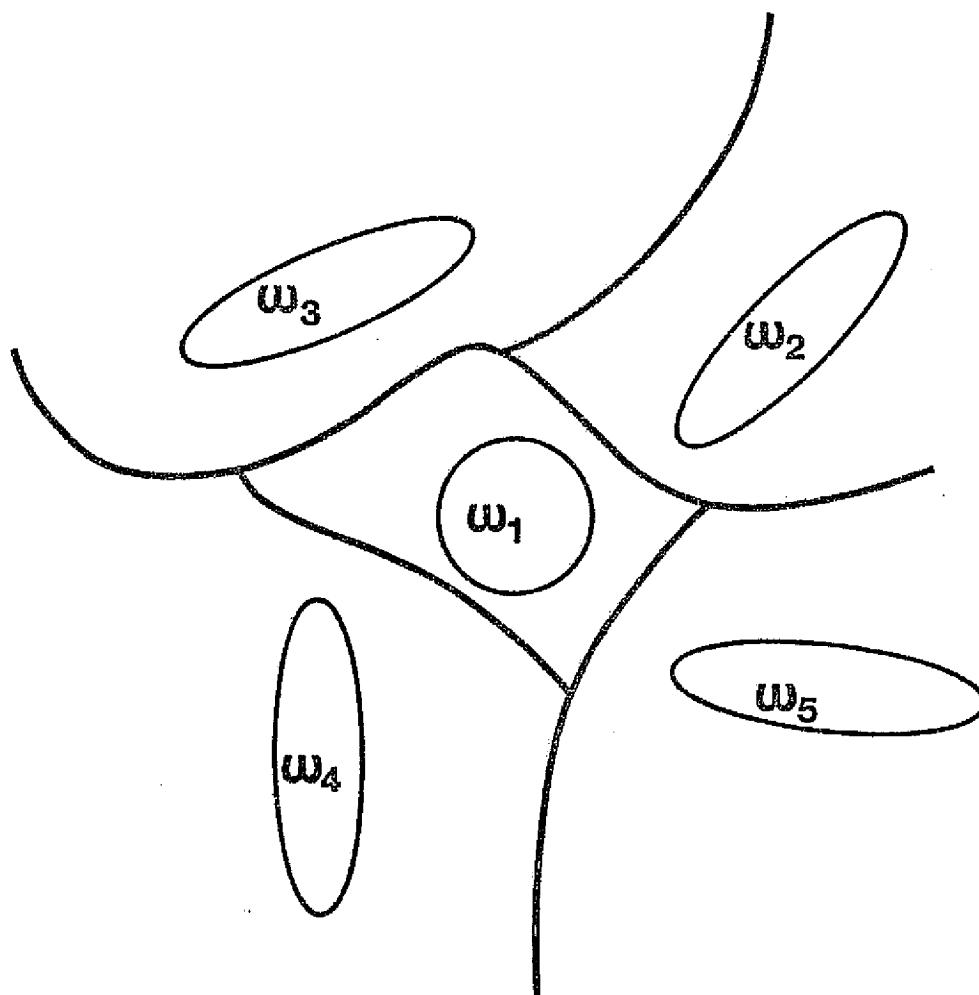


Fig. 2-3 Optimal Partitioning of the Feature Space and the Resulting Quadratic Boundaries.

In this section, the difficulties associated with a direct evaluation of the multidimensional classification error integral were discussed. The parameters of the problem, in order of decreasing contribution to the problem complexity, are listed below

1. Inequality of covariance matrices
2. Dimensionality of feature space
3. Multiplicity of populations

2.3 Approximation to the Classification Error Integral

In the previous section the Bayes error was expressed as a multiple integral over R^N , the N dimensional Cartesian coordinate system of the feature space. The underlying difficulties in evaluation of (2-6) were attributed to the intractable mathematical description of the contours of Γ_i , and the N -th order multiple integral over an arbitrary shaped domain, Γ_i . There are two transformations that would circumvent these problems.

2.3.1 Coordinate Transformation

The multispectral scanner detects the reflected electromagnetic energy in a number of optical and infrared bands. Although these bands are essentially non-overlapping, the responses observed are correlated. A rise in signal amplitude in one band is accompanied by a similar effect in an adjacent band. In statistical terms this property translates into a probability space with correlated variates.

The M populations are represented by a set of general mean and covariance matrices. Great algebraic simplification would occur if every $\underline{\Sigma}$ was in a diagonal form resulting in separable density functions. This simplification stems from the application of product rule. Let $\underline{X} = (x_1, x_2, \dots, x_N) \in B$ and $\underline{Y} = (y_1, y_2, \dots, y_{N'}) \in G$ be in Euclidean spaces of N and N' dimension respectively. A Cartesian Product BxG, is a space of N+N' dimensions with points $(x_1, x_2, \dots, x_N, y_1, y_2, \dots, y_{N'})$ such that $(x_1, x_2, \dots, x_N) \in B$ and $(y_1, y_2, \dots, y_{N'}) \in G$. Let there exist an m point integration rule, R, over B

$$R(f) = \sum_{i=1}^m a_i f(\underline{X}_i) \approx \int_B f(\underline{X}) dV \quad \underline{X}_i \in B \quad (2-12)$$

and an n point integration rule, R', over G

$$R'(f) = \sum_{j=1}^n b_j f(\underline{Y}_j) \approx \int_G f(\underline{Y}) dV \quad \underline{Y}_j \in G \quad (2-13)$$

Then the product rule of R and R' defined over BxG is given by

$$R \times R' = \sum_{i=1}^m \sum_{j=1}^n a_i b_j f(\underline{X}_i, \underline{Y}_j) \approx \int_{B \times G} f(\underline{X}, \underline{Y}) dV \quad (2-14)$$

From these properties, it quickly follows that if R integrates $f(\underline{X})$ exactly over B and R' integrates $g(\underline{Y})$ exactly over G then provided $h(X,Y) = f(X)g(Y)$, $R \times R'$ integrate $h(X,Y)$ exactly over BxG. A brief proof of this theorem given in [47] follows.

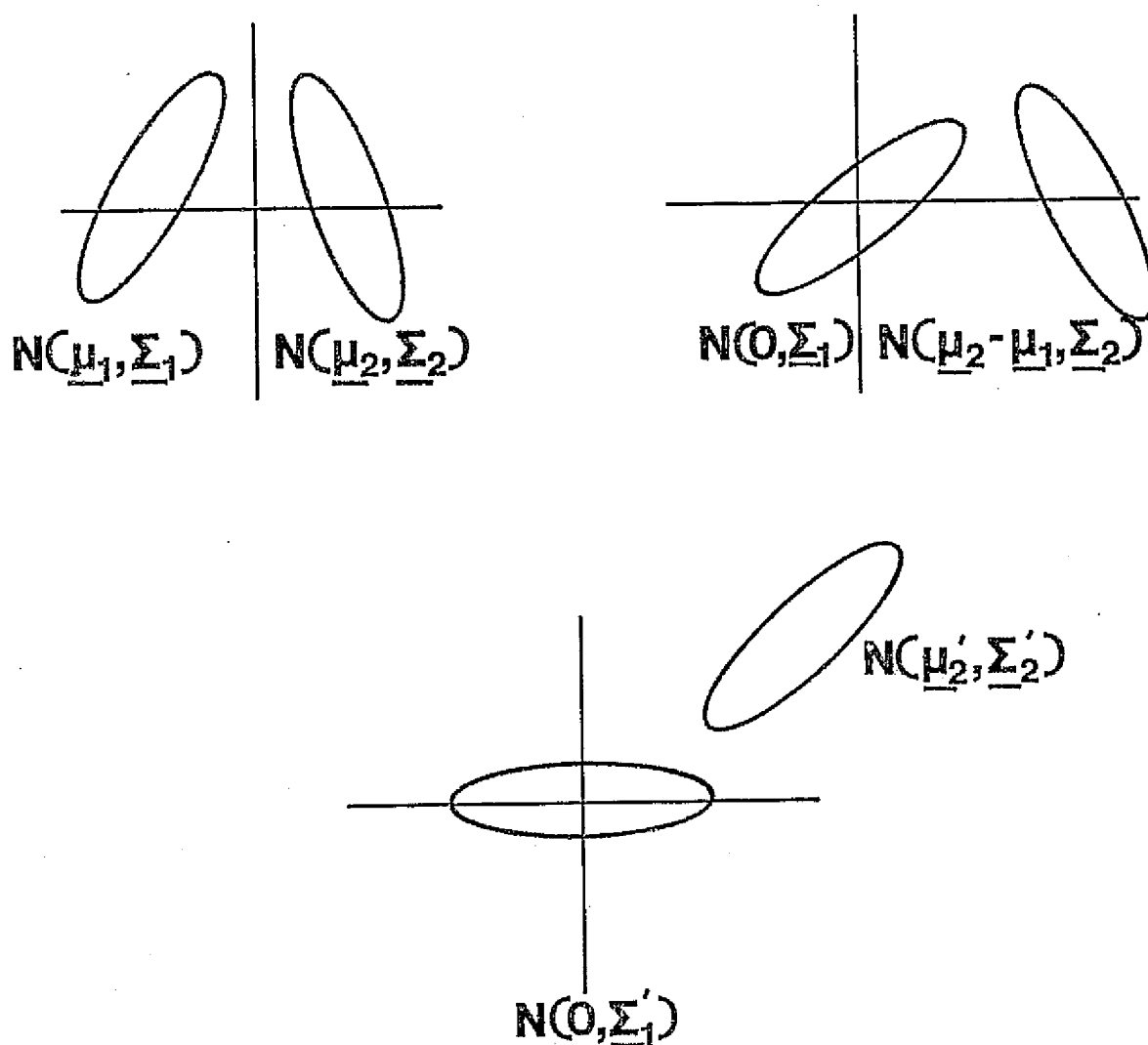


Fig. 2-4 Three Linear Transformation Steps Prior to Feature Space Sampling.

ORIGINAL PAGE IS
OF POOR QUALITY

$$\begin{aligned}
\int_{B \times G} h(\underline{X}, \underline{Y}) dV &= \int_{B \times G} f(\underline{X}) g(\underline{Y}) dV = \int_B f(\underline{X}) dV_B \int_G g(\underline{Y}) dV_G \\
&= \sum_{i=1}^m a_i f(\underline{X}_i) \sum_{j=1}^n b_j g(\underline{Y}_j) \\
&= \sum_{i=1}^m \sum_{j=1}^n a_i b_j f(\underline{X}_i) g(\underline{Y}_j) \\
&= R \times R'
\end{aligned} \tag{2-15}$$

Potentially, this rule can reduce the dimensionality of a problem from N to 1. Such a property is not an intrinsic feature of the remotely sensed data, however. Moreover, there is no transformed space in which $M(M>2)$ covariance matrices can be represented in diagonal form.

Since the calculation of $\hat{P}_{c|\omega_i}$ precedes the estimation of overall classification accuracy, an M stage successive estimation procedure in M linearly related probability spaces can be envisioned. For example, stage i consists of the following mapping

$$\left. \begin{aligned} \underline{\mu}'_j &= \underline{\mu}_j - \underline{\mu}_i \\ \underline{M}_j &= \phi^T \underline{\mu}'_j \\ \underline{S}_j &= \phi^T \underline{\Sigma}_j \phi \end{aligned} \right\} j=1, 2, \dots, M \tag{2-16}$$

where ϕ is the eigenvector matrix derived from $\underline{\Sigma}_i$. Therefore, in each transformed space, $T_i(\Omega)$, ω_i has a null mean vector and a diagonal covariance matrix. Fig. 2-4 is a pictorial representation of (2-16) for two classes. This unitary transformation is linear, preserves the Euclidean distance and

pairwise divergence and the probability of error is invariant under such mapping. It will be shown that formulation of $\hat{P}_c|_{\omega_i}$ in $T_i(\Omega)$ will provide an N to 1 dimensionality reduction.

2.3.2 Discrete Space Approach

For any continuous formulation of a problem there exists a discrete counterpart, specific choice of which is dependent upon individual cases and requirements. Let Ω be the continuous probability space. A transformation, T , is required such that in $T(\Omega)$, Γ_i can be completely described in a non-parametric form, thereby bypassing the requirement for an algebraic representation of Γ_i . This desired transformation would sample Ω into a grid of N -dimensional cells according to a certain rule; thus, expressing the Bayes error integral in the discrete space of $T(\Omega)$.

The sampling of the probability space is equivalent to the discrete representation of the random variates along each feature axis. The multispectral data is generally modeled as a multivariate normal random process. What is required, therefore, is a discrete approximation to a normal random variable that would exhibit desirable limiting properties. Let $y_n \sim \text{Bi}(n, p)$ be a binomial random variable with parameters n and p . Then x_n defined by

$$x_n = \frac{y_n - np}{\sqrt{np(1-p)}} \quad y_n = 0, 1, 2, \dots, n \quad (2-17)$$

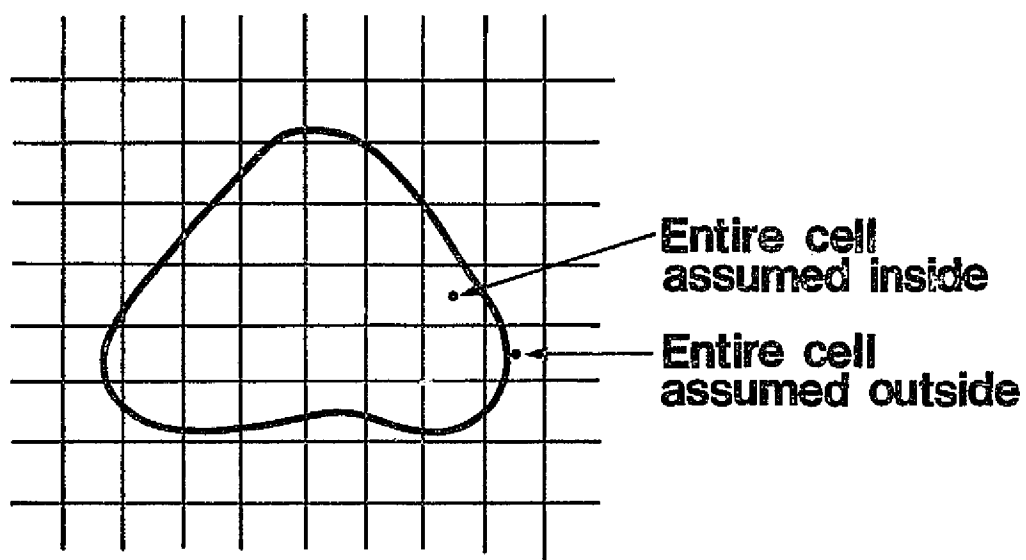


Fig. 2-5 Boundary Location Estimation.

converges to $x \sim N(0,1)$ in distribution [48]; i.e.,

$$\lim_{n \rightarrow \infty} F_n(x) \rightarrow F(x)$$

The convergence is most rapid if $p = \frac{1}{2}$. Then

$$x_n = \frac{(Y_n - n/2)^2}{\sqrt{n}} \quad (2-18)$$

The variance of x_n is set equal to the eigenvalue of the transformed Σ_i by incorporating a multiplicative factor $(\sqrt{\lambda_i})$ in (2-18).

The segmentation of Ω by a union of elementary hyper-volumes makes nonparametric representation of Γ_i and its contours feasible. Some comments regarding the structure of the sampled Ω are in order. The coordinates of each cell's center are known and given by (2-18). The spacing between the centers is readily shown to be equal to $\delta_i = \frac{2}{\sqrt{n}} \sigma_i$ along the i th axis. The grid extent is therefore $\pm \sqrt{n} \sigma_i$ with $n+1$ cells along each coordinate axis. The simultaneous solution of the set of M N th degree polynomials is now reduced to the identification of each cell with one of M partitions within Ω .

Specifically, following the orthonormal transformation on ω_i and sampling of Ω accordingly, each cell's coordinate is assigned to the appropriate Γ . This process is carried on exhaustively, therefore Γ_i can be defined as a set such that

$$\Gamma_i = \{UX_{-n} : X_{-n} \in \Gamma_i\} \quad (2-19)$$

Fig. 2-5 shows a pictorial representation of (2-19). The description of domains of integration as a union of elementary units alleviates the need for the precise knowledge of the

boundary location, although the sampling grid can provide an estimate within one δ . Once the exhaustive process of assignment is completed $\hat{P}_{c|\omega_i}$, the integral of $f(\underline{X}|\omega_i)$ over Γ_i , is represented by the sum of hypervolumes over the elementary cells within Γ_i .

Using this procedure, cumbersome implementation of numerical integration techniques in multidimensions are avoided. One of the main features of the orthonormal transformations preceding the sampling process is the decoupling of Σ_i , thereby generating the separability property of the transformed $f(\underline{X}|\omega_i)$ along each dimension. Invoking the product rule and designating the domain of a cell, centered at the origin within Γ_i , as C_i :

$$\int_{C_i} f(\underline{X}|\omega_i) d\underline{x} = \int_{-\frac{\delta_1}{2}}^{\frac{\delta_1}{2}} f(x_1|\omega_i) dx_1 \int_{-\frac{\delta_2}{2}}^{\frac{\delta_2}{2}} f(x_2|\omega_i) dx_2 \dots \int_{-\frac{\delta_N}{2}}^{\frac{\delta_N}{2}} f(x_N|\omega_i) dx_N \quad (2-20)$$

This unit of probability volume is equal to the product of N one dimensional normal integrals, the value of which is widely tabulated. Thus, no involved numerical procedure is required to evaluate (2-20).

The relationship expressed in (2-20) is the building block in the probability of error estimation. Referring to this algorithm as a 'Controlled Space Partitioning' (CSP) we can write the conditional classification accuracy estimate as

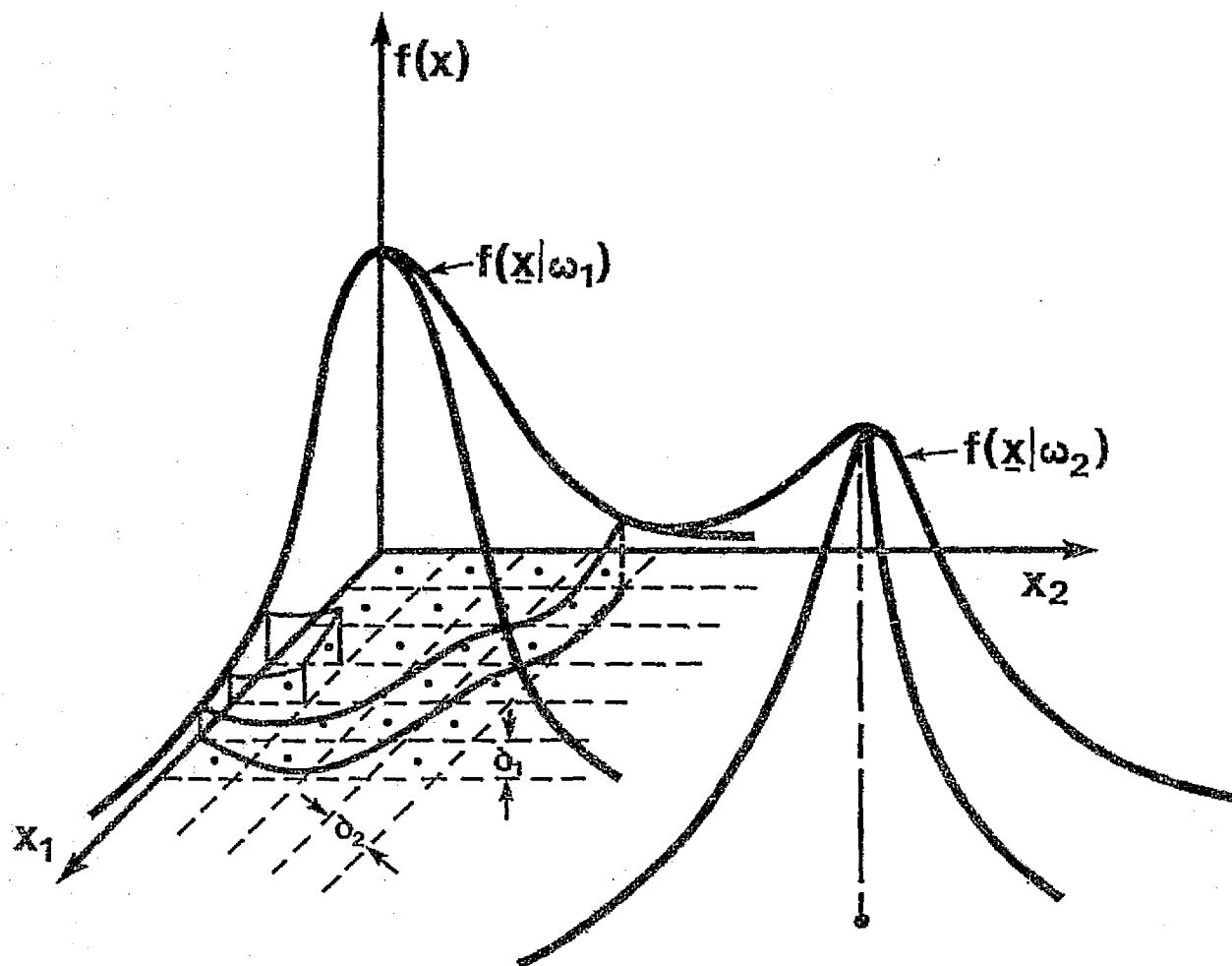


Fig. 2-6 A Conceptual Illustration of the CSP Error Estimation Technique Using Two Features.

$$\hat{P}_C | \omega_i = \sum_{C \in \Omega} \int_{c_1 - \frac{\delta_1}{2}}^{c_1 + \frac{\delta_1}{2}} f(x_1 | \omega_i) I_i(C) dx_1 \int_{c_2 - \frac{\delta_2}{2}}^{c_2 + \frac{\delta_2}{2}} f(x_2 | \omega_i) I_i(C) dx_2$$

$$\dots \int_{c_N - \frac{\delta_N}{2}}^{c_N + \frac{\delta_N}{2}} f(x_N | \omega_i) I_i(C) dx_N \quad (2-21)$$

$$\hat{P}_C = \sum_{i=1}^n P(\omega_i) \hat{P}_C | \omega_i$$

where
$$I_i(C) = \begin{cases} 1 & \text{if } C \in \Gamma_i \\ 0 & \text{otherwise} \end{cases}$$

C = The domain of an elementary cell

Fig. 2-6 is a geometrical representation of (2-21).

2.4 Error Analysis

Formulation of a problem with inherently continuous parameters in a discrete space as a means of approximation or estimation of the end product necessarily incurs errors that need to be studied. Error terms cannot be expressed in the form of exact expressions and can only be bounded or put in some defined statistical model; otherwise, the approximation would be exact. Extension of the one dimensional integration error analysis results does not appear to be possible due to the lack of any correspondence between the unidimensional and multidimensional integration domains. In the multivariate

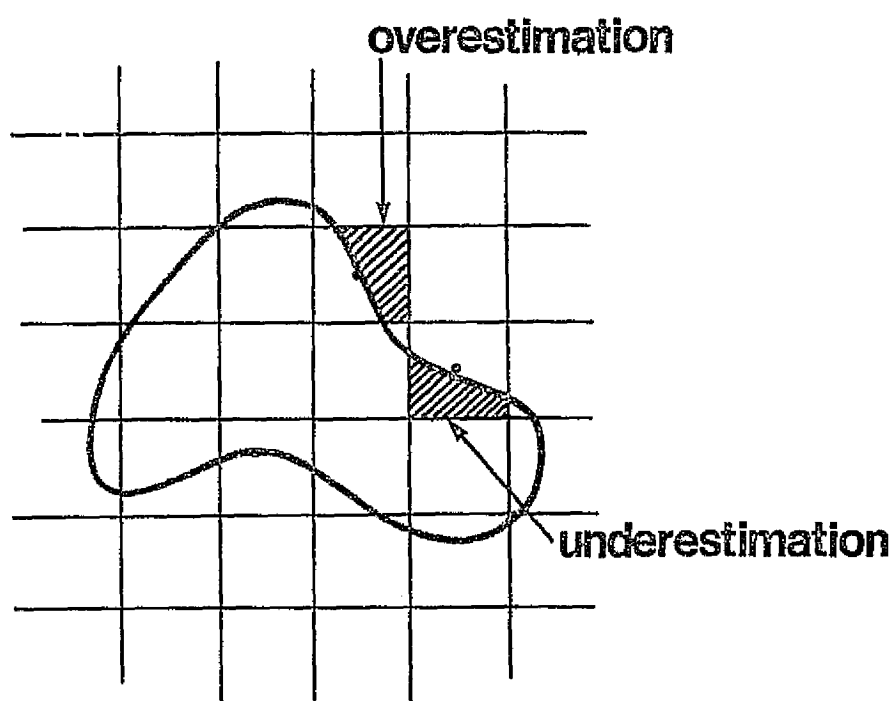


Fig. 2-7 Type II Error Structure.

integration field, the errors studied are related to the simple domains such as hypercubes, simplexes etc. [49].

There are basically two types of error encountered in the implementation of the CSP algorithm. Type I error originates in the one dimensional integration of a normal density function over the region (a,b) . This quantity is available both in tabular form and as FORTRAN callable subroutine subprograms which are capable of supplying arbitrarily high accuracy results. Type II error occurs only at the boundary of r_i because the sampled grid essentially estimates the location of such contours. A self cancelling property of this type of error is brought about by the geometrical structure of the regions when: (a), $f(X|\omega_i)$ is integrated over the whole elementary hypercube instead of a portion inside r_i (grid point $\underline{x}_n$ close to the boundary and $\underline{x}_n \notin r_i$); and (b), $f(X|\omega_i)$ is not integrated over a portion of r_i (grid point $\underline{x}_n$ close to the boundary but $\underline{x}_n \notin r_i$). (a) adds a positive bias and (b) adds a negative bias to the result of integration, $\hat{p}_c$. For a sampling grid with fine subdivision and over the ensemble location of all the boundaries, the events

$$\{\underline{x}_n \in r_i \text{ or } \underline{x}_n \notin r_i | \underline{x}_n \text{ near the boundary}\} \quad (2-22)$$

have equal likelihood; hence, positive and negative biases occur equally often. Fig. 2-7 shows the structure of type II error in 2-dimensions.

2.4.1 Statistical Properties of the Estimate

The error encountered in estimating the classification error is primarily of type II. Much insight into the structure

of this error is obtained by examining the problem in one dimension. Let $f(x|\omega)$ be a class conditional Gaussian density function, $N(0,1)$, and let x_b be a fictitious unknown boundary possibly separating ω from some other population. A grid of size n is set up and it is determined that $x_{n_0} \in \Gamma_i$ and $x_{n_0+1} \notin \Gamma_i$

Fig. 2-8. x_{n_0} has equal likelihood of being inside or outside Γ_i . Equivalently, it can be stated that x_b is random with uniform variations within one δ , i.e.,

$$x_b \sim U(x_{n_0} - \frac{\delta}{2}, x_{n_0} + \frac{\delta}{2}) \quad (2-23)$$

The error in estimating the area of Γ_i can therefore be represented as

$$e = \int_{x_{n_0}}^{x_b} f(x) dx \quad (2-24)$$

which, depending on x_b , can take on either positive or negative values. The expected value of e is

$$\begin{aligned} \bar{e} &= \int_{x_{n_0} - \frac{\delta}{2}}^{x_{n_0} + \frac{\delta}{2}} \left[\int_{x_{n_0}}^{x_b} f(x) dx \right] f(x_b) dx_b \\ &= \frac{1}{\delta} \int_{x_{n_0} - \frac{\delta}{2}}^{x_{n_0} + \frac{\delta}{2}} (Q(x_{n_0}) - Q(x_b)) dx_b \\ &= Q(x_{n_0}) - \frac{1}{\delta} \int_{x_{n_0} - \frac{\delta}{2}}^{x_{n_0} + \frac{\delta}{2}} Q(x_b) dx_b \end{aligned} \quad (2-25)$$

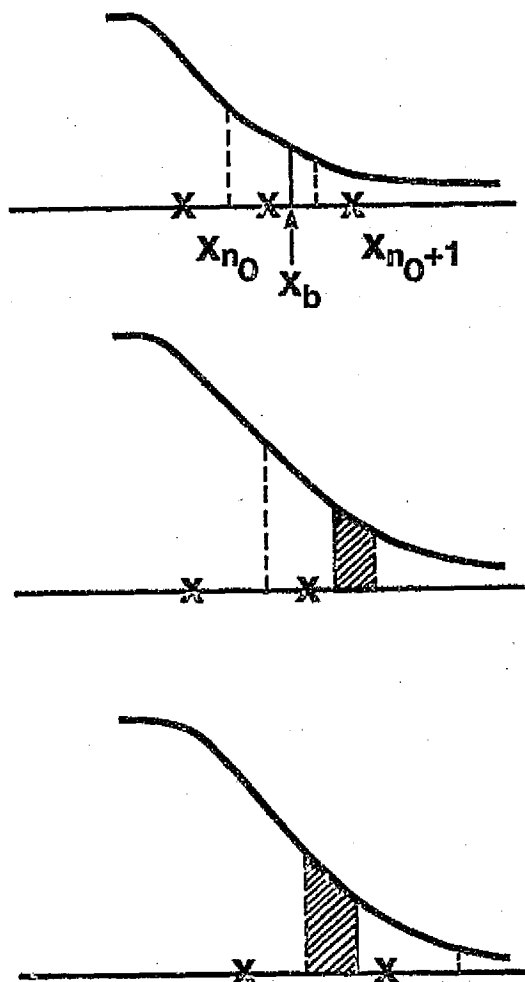


Fig. 2-8 A Case of Net Negative Estimator's Bias.

The variation of $\bar{e}$ vs. x_{n_0} is plotted in Fig. 2-9.

Examination of $\bar{e}$ shows that although small, it is mostly negative. Its magnitude decreases with increasing n and x_{n_0} . These properties can also be deduced geometrically. The negative bias is due to the following obvious inequality

$$Q(a) - Q(b) > Q(c) - Q(d)$$

given that

$$\begin{cases} a < b < c \\ c - a > 0 \\ d - b > 0 \\ a, b, c, d \geq 0 \end{cases} \quad (2-26)$$

Fig. 2-8 shows two cases where the closest cell to the boundary is considered either inside ($x_{n_0} \in \Gamma_i$) or outside ($x_{n_0} \notin \Gamma_i$) the decision region resulting in an over and underestimation of correct classification, respectively. From (2-26) it then follows that the magnitude of negative bias is greater than that of positive bias. Thus, this procedure gives estimator with a net negative bias.

A different situation exists when the region of integration is doubly connected as in Fig. 2-10. In this case the shift of a cell center from just inside the boundary to just outside, produces an opposite effect. Whereas in the previous case such a shift would have reversed the sign of the bias term from positive to negative with an increase in magnitude, in the new domain the net change in bias will be positive simply because the inside-to-outside move now is

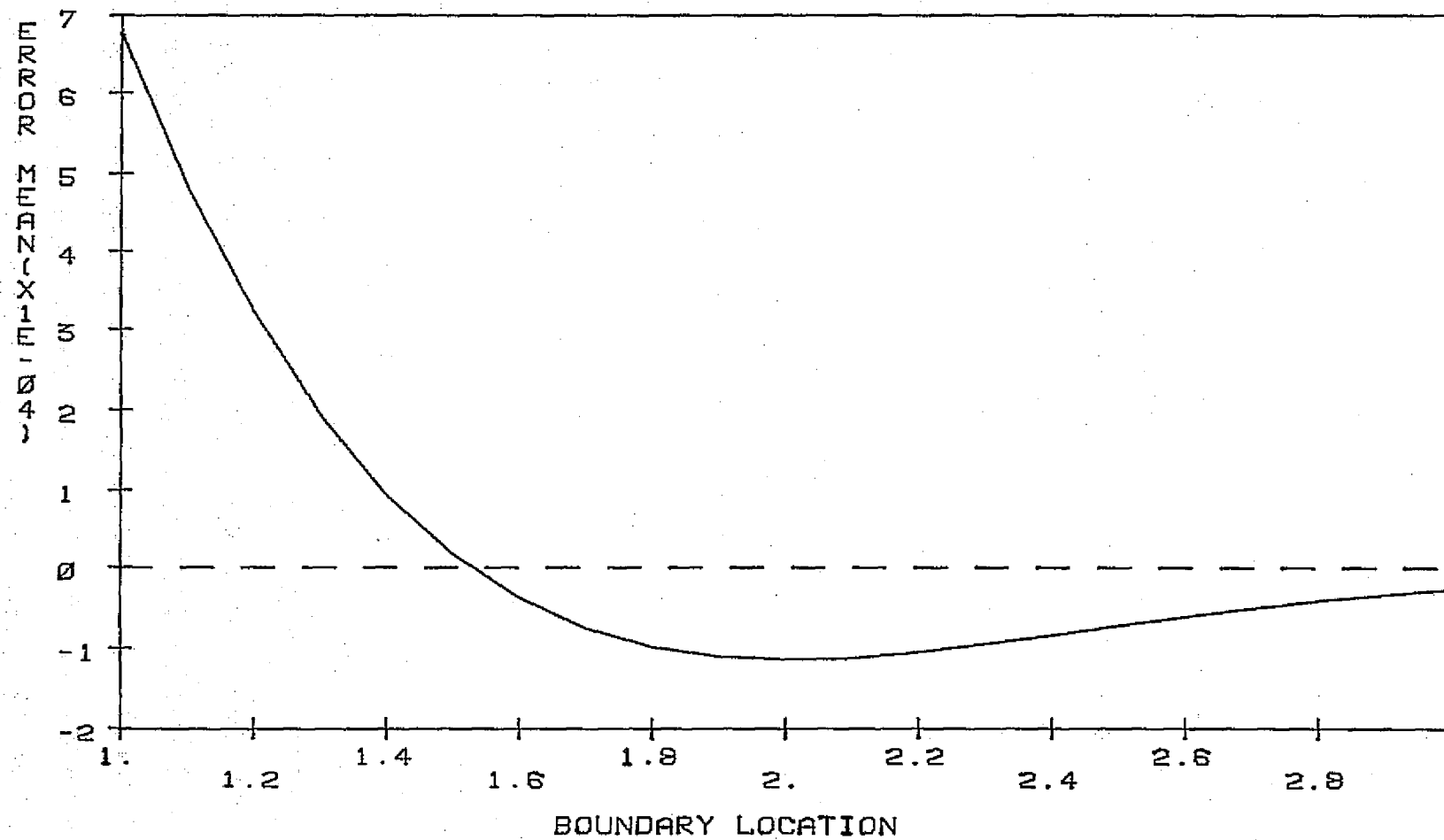


FIG. 2-9 EXPECTED VALUE OF AN AREA ESTIMATOR USING
EQUIDISTANT SAMPLES. 64 CELLS PER AXIS.

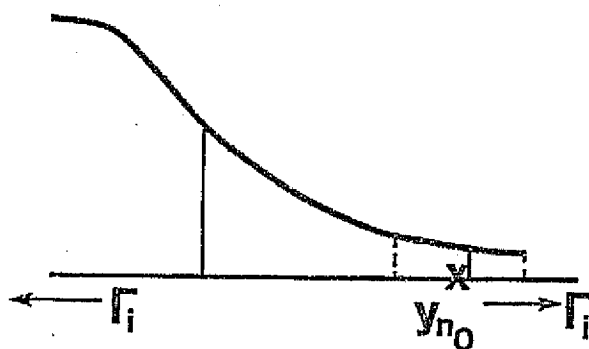
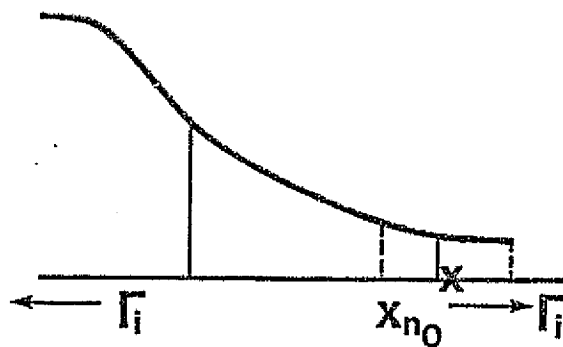


Fig. 2-10 A Case of Net Positive Estimator's Bias.

toward the mean rather than away from it.

In N dimensions the total error in estimating the conditional probability of correct classification can be represented by a weighted sum of the boundary errors,

$$e_T = \sum_{i=1}^{N_B} w_i e_i \quad (2-27)$$

where N_B is the number of cells along the boundary, w_i is the weighting sequence and e_i is the N dimensional error associated with the i th boundary cell. In order to obtain a variance expression for e_T , the statistical properties of e_i need to be determined. From (2-22) it follows that the location of the boundary is uniformly distributed within one boundary cell width. In general, it does not follow that the volume error is also uniformly distributed within one cell volume. This is strictly true only in cases where the decision boundary and the boundary of a cell are 'parallel'. Adoption of a uniform distribution assumption for e_i , however, provides a considerable simplification in the derivation of an expression for the variance of the error. With regard to the first expression for the variance of e_T it should be noted that the assumption of a uniformly distributed e_i generates a variance higher than the true value. Thus, the resulting expression can be taken as an upper bound on the variance of e_T . Let

$$e_i \sim U\left(-\frac{V_C}{2}, \frac{V_C}{2}\right) \quad (2-28)$$

ORIGINAL PAGE IS
OF POOR QUALITY

where v_c is the volume of an elementary cell given by

$$v_c = \prod_{i=1}^N \delta_i \quad (2-29)$$

The contribution of e_i to the total integration error should clearly be weighted to incorporate the effect of boundary location relative to the mean of ω_i . An appropriate choice for the weighting sequence w_i is the height of $f(\underline{X}|\omega_i)$ at a particular boundary point. A weighting process such as this effectively assigns a 'significance' to each e_i . Although the magnitude of e_i may have been large in the context of volume approximation of Γ_i , if the normalized distance of such cell to μ_i is large it generates negligible volume under $f(\underline{X}|\omega_i)$.

In order to obtain a variance expression for e_T the small bias is assumed negligible. The variance of individual errors, e_i is

$$\text{Var}\{e_i\} = \frac{v_c^2}{12} = \alpha_i^2 \quad (2-30)$$

Therefore,

$$\begin{aligned} \text{Var}\{e_T\} &= E\{e_T^2\} = E\left\{ \sum_{i=1}^{N_B} \sum_{j=1}^{N_B} w_i w_j e_i e_j \right\} \\ &= \sum_{i=1}^{N_B} w_i^2 \overline{e_i^2} + \sum_{i=1}^{N_B} \sum_{\substack{j=1 \\ i \neq j}}^{N_B} w_i w_j \overline{e_i e_j} \\ &= \sum_{i=1}^{N_B} w_i^2 \alpha_i^2 + \sum_{i=1}^{N_B} \sum_{\substack{j=1 \\ i \neq j}}^{N_B} w_i w_j \rho_{ij} \sigma_i \sigma_j \end{aligned} \quad (2-31)$$

where ρ_{ij} is the correlation coefficient between e_i and e_j . Obtaining an analytical expression for ρ_{ij} in N dimensions is

theoretically feasible however its complexity considerably diminishes the usefulness of (2-31). Because the ρ_{ij} are small for widely separated boundary cells, a reasonable approximation to the $\text{Var}\{e_T\}$ is given by

$$\text{Var}\{e_T\} = \sum_{i=1}^{N_B} w_i^2 \alpha_i^2 \quad (2-32)$$

Expanding (2-32)

$$\alpha_i^2 = \frac{1}{12} \left(\prod_{i=1}^N \delta_i \right)^2 \quad (2-33)$$

$$w_j = \frac{1}{(2\pi)^{N/2} \prod_{i=1}^N \sigma_i} e^{-\frac{1}{2} \mathbf{X}_j^T \Sigma^{-1} \mathbf{X}_j} \quad \mathbf{X}_j \in \Gamma_{B_i} \quad (2-34)$$

where Γ_{B_i} is the boundary domain of Γ_i . Substituting

$$\delta_i = \frac{2}{\sqrt{n}} \sigma_i \text{ in (2-33)}$$

$$\begin{aligned} \alpha_i^2 &= \frac{1}{12} \left(\prod_{i=1}^N \frac{2}{\sqrt{n}} \sigma_i \right)^2 \\ &= \frac{2^{2N}}{12n^N} \prod_{i=1}^N \sigma_i^2 \end{aligned} \quad (2-35)$$

Therefore, (2-32) is equal to

$$\text{Var}\{e_T\} = \frac{1}{12} \left(\frac{2}{n\pi} \right)^N \sum_{i=1}^{N_B} f_i^2 \quad (2-36)$$

where f_i is the exponential part of (2-34).

ORIGINAL PAGE IS
OF POOR QUALITY

The variation of $\text{Var}\{e_T\}$ vs. n and N cannot be fully explored due to the $\sum f_i^2$ factor which is problem-dependent. However, it does follow that P_C is convergent in the mean-square sense;

$$E\{|P_C - P_C|^2\} = \text{Var}\{e_T\} \rightarrow 0 \text{ for } n \rightarrow \infty \quad (2-37)$$

This observation is less than obvious since as n increases so do N_B and $\sum f_i^2$ and, hence, could potentially be self-canceling. Although the increase of N_B and n are monotonic, experimental evidence suggests that N_B as a percentage of cells within r_i steadily decreases so while more cells are allocated to r_i , comparatively fewer ones reside near the boundary. Therefore, $\sum f_i^2$ only slows the convergence of the variance towards zero. It also follows from (2-37) that variance decreases for high recognition rates (i.e., small f_i).

2.5 General Comments

Formulation of a problem in an N -dimensional space requires coping with situations not present in the single dimension case. In addition to the mathematical complexities, the practicality of implementation of any method should be closely examined. In particular, with the digital computer capability and its cost as the ultimate limiting factor, the computation time of processing in an N -dimensional domain takes on a prime importance.

Techniques requiring exhaustive enumerations can be potentially expensive, in many instances totally beyond the

available resources. The implementation of (2-21) requires processing of an N-dimensional grid of points. With $(n+1)$ cells along each axis, there are $(n+1)^N$ such points to be allocated to their respective domains. In one dimension, very accurate estimates can be obtained long before the size of $n+1$ presents any computational difficulties. The multi-dimensional case is different. The exponential rise of the grid size with N can make the execution time prohibitively long. This 'dimensionality effect' can effectively generate a computational barrier and thus render the algorithm inoperative if n is 'too large'.

The quality of the estimate as shown in sec. 2.4 is dependent on the size of n; i.e., the grid fineness. So the central question is whether n can be large enough to generate a high quality estimate and yet small enough to make the estimation process feasible. The sampling grid, the algorithm and remotely sensed data itself have properties that help answer this question in the affirmative. The current MSS system in operation collects data in four spectral bands. It is believed that future space platforms primarily Landsat C will be equipped with scanners having data collection capability not beyond five spectral bands. Therefore, N for all useful purposes is limited to the 4 to 6 range. Actually, optimum processes may not utilize all of these bands due to their redundancy.

The next question is the relative magnitude of n. The answer lies partly in the outer location of the desired

boundaries and the fact that the sampling grid must cover the entire relevant domains. It turns out that in most cases of interest, Γ_i is a simply connected domain but Z_i is a doubly connected domain. This approach toward the evaluation of the classification performance through the estimation of probability of correct classification ensures that sampling of the probability space is confined to a closed finite domain of Γ_i , thus alleviating the need to sample Z_i , a far larger region. Having established that Γ_i is bounded in many cases, the question now is whether the outer limits of the grid will encompass the appropriate boundary and thus sample Γ_i thoroughly with a reasonable magnitude for n .

Define g_{e_i} as the extent of the sampling grid along the i th transformed feature axis. From sec. 2.3.2, $|g_{e_i}| = \sqrt{n}\sigma_i$. Although no such quantity can be precisely defined for Γ_i , let r_{e_i} be the outer limits of Γ_i in some average sense. Two cases can be distinguished: (a), $r_{e_i} < |g_{e_i}|$ in which case clearly, the grid has sampled the entire domain of interest; and (b), $r_{e_i} > |g_{e_i}|$, a condition which either means that n is exceedingly small or that r_{e_i} is located very many σ_i 's away from μ_i . In this case any error committed but unaccounted for, will have very small effect on the outcome due to the negligible volume under a normal density function for anything more than a few standard deviations from the origin. Since in most applications $n \geq 8$, the grid extent will be $> \pm 2.2\sigma$ and will satisfactorily sample the entire domain of interest.

CHAPTER 3

Line Scanner Imaging Systems

The primary goal of a remote sensing system is the collection from a scene of reflected or emitted electromagnetic energy in selected spectral bands. This task has been traditionally accomplished by airborne photographic equipment and is analyzed by photointerpretation techniques. There are several major drawbacks associated with such a method.

The sensitivity of photographic films is generally limited to the near ultraviolet-near infrared band; therefore, night time operation is severely limited unless the scene is externally illuminated. Clouds, fog and smoke are opaque through this portion of electromagnetic spectrum. Most importantly, handling of the film itself is awkward and the accompanying telemetry problem can make its deployment aboard a nonrecoverable vehicle unattractive.

Nonphotographic sensors overcome many of these shortcomings. Through the selection of the proper detector, spectral coverage can be extended to microwave and beyond where clouds and bad weather do not seriously hinder the sensor's performance. Having the data in the form of an electrical signal lends itself to efficient and powerful

transmission and processing techniques.

3.1 Types of Systems

The majority of current remotely sensed data is obtained in the ultraviolet, visible and infrared portions of the spectrum by scanning systems. One of the earliest of such systems was 'Reconofax' operating in the visible region [50]. It used either moonlight or an internal illumination source to produce maps of the ground scene at night. Lack of detectors with rapid rise time, produced imagery with unsatisfactory resolution compared to photographic methods. As a result of improvements in detector technology, current scanning system can produce imagery of high quality within a reliable, compact and fairly simple structure.

3.1.1 Multispectral Scanners

A widely used earth resources data gathering system is the electro-optical scanning radiometer otherwise known as a multispectral scanner. A MSS is generally an object plane scanner [51] and consists of a rotating mirror and a telescope that directs reflected energy from a small portion of the object plane. A bank of detectors responding to different wavelengths receives the incoming radiation which, after detection, sampling and quantization, is telemetered to the ground station. When such a system is placed

in an aircraft or an earth orbiting satellite a strip map of the ground scene is produced. The cross-track coverage is performed by the oscillating mirror and vehicle motion accomplishes the along-track coverage. Contiguous coverage is required to prevent underlap. This can occur if the satellite speed is too high or the mirror's rotational rate too slow.

This simple structure can be upgraded to include the currently employed scanners in which an n sided mirror rotates at a rate of r revolution per second thereby producing n lines for each rotation. There are a total of d detectors and, thus, d lines are scanned by each side of the mirror. A total of nxd lines are scanned for a full rotation.

Let k_τ be the dwell time of a detector on each resolution element and V and H be the speed and altitude of the vehicle, respectively. It can be shown [50] that subject to a dwell time not less than k_τ and a no underlap scanning mode, the angular resolution of an MSS has a lower bound given by

$$\phi \geq (2\pi k/nd) (V/H) (\tau) \quad (3-1)$$

with equality for contiguous lines. From a hardware point of view, the adjustable parameters are limited. V and H are interdependent and are determined by orbit considerations. τ is a property of the detector. n and d are variable

parameters to choose as a means of the MSS instantaneous field of view (IFOV) control.

One of the most widely used operational remote sensing instruments is the Landsat multispectral scanner. Landsat, an Earth resources monitoring satellite, is positioned on a polar orbit at an altitude of about 900 km with a complete global coverage cycle of 18 days. The vehicle's operation is chosen so as to provide a 14% scan overlap. The MSS collects data in 4 spectral bands, two visible and two near infrared, all in spatial registration. Six lines are scanned simultaneously and with an IFOV of $87 \mu\text{rad}$ providing a ground resolution of about 80 m, with a total cross-range width of 185 km. The Skylab S192 scanner provided similar resolution with 13 spectral bands from $0.5 \mu\text{m}$ to $12.5 \mu\text{m}$. Among other MSS systems is the Thematic Mapper for Landsat D. Spectral coverage is extended to 7 bands from $.51 \mu\text{m}$ to $2.35 \mu\text{m}$ with some gaps plus a thermal band from $10.4 \mu\text{m}$ to $12.6 \mu\text{m}$. Angular resolution of $33 \mu\text{rad}$ will correspond to a ground IFOV of 30 m at a 900 km altitude [52].

3.2 System Modeling of a Multispectral Scanner System

The objectives outlined in the introductory chapter required a parametric representation and evaluation of the MSS performance. Like any other complex and integrated system, the multiplicity of parameters is numerous. Sensor

choice, band selection and telemetry links, are but a few of the interacting components of the system design. From the viewpoint of information extraction and processing, however, the spatial characteristics of a scanner along with the spatial resolution and additive noise take on a particular significance.

Modeling of the MSS by a linear system opens the way to the application of existing techniques in system theory. Since the classification accuracy is totally a function of class statistics under the Bayes rule, examination of the random process transformation carried out by the scanner PSF can be most revealing.

Topics of particular interest are

1. Effect of the scanner IFOV on population statistics.
2. Effect of data spatial correlation on the classification accuracy.
3. Effect of signal-to-noise ratio on classification accuracy.
4. Trade off between spatial resolution and SNR.
5. Effect of spatial resolution on classification accuracy.
6. The interactive relationship between IFOV, spatial correlation, class statistics, SNR and classification accuracy.

3.2.1 MSS Spatial Model

The incident electromagnetic energy after reflection from a target is detected by the scanner IFOV. The ultimate goal of such operation is a perfect reproduction of the radiant energy. This objective cannot be accomplished with any physically realizable system. Finite IFOV, required by detector sensitivity among other things, keeps the ground resolution at a finite level. The resolution degradation can be subsequently dealt with through various image enhancement techniques [53,54].

The averaging operation performed by the scanner point spread function can be modeled by a linear shift-invariant multiple-input, multiple-output system. Input signals consist of N random processes in N spectral bands corrupted by atmospheric noise and scattering. Each input is linearly transformed by the scanner PSF and additional detector and pre-amp noise further contributes to the signal degradation.

Fig. 3-1 is a basic block diagram of this spatial model. $h(x,y)$ is the two dimensional PSF to be specified for any desired system. In particular where the MSS is concerned, the assumption of a Gaussian shaped IFOV has been widespread. The justification for this is essentially satisfactory experimental results and perhaps equally important is the mathematical convenience of this model. Note that the results obtained hereafter are fundamentally

ORIGINAL PAGE IS
OF POOR QUALITY

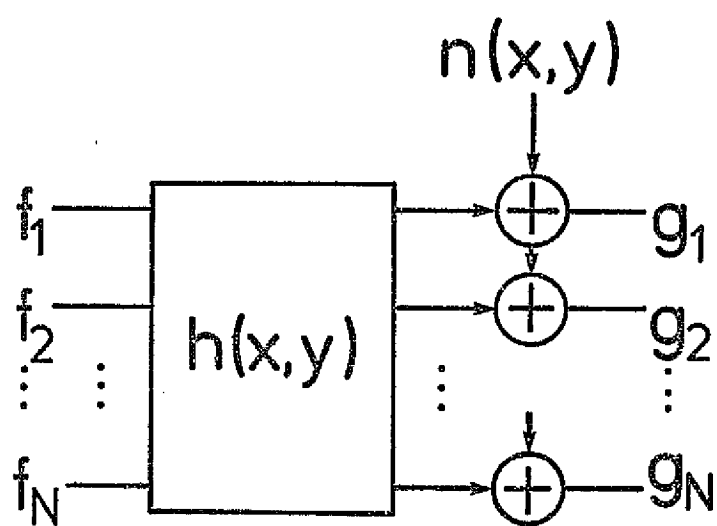


Fig. 3-1 MSS Spatial Model as a Linear System.

independent of the functional form of the PSF. However, using this assumption, it is frequently possible to obtain closed form expressions and to make comparisons with alternate methods a majority of which adhere to the same assumption.

In a two dimensional plane a Gaussian PSF is specified by the following relationship

$$h(x,y) = c_1 e^{-\frac{x^2}{r_0^2}} e^{-\frac{y^2}{r_0^2}} \quad (3-2)$$

The important parameter is r_0 , PSF's characteristic length, which in effect determines the ultimate ground resolution and noise content of the collected data. Increasing r_0 results in a deterioration of the former but improvement of the latter. The significant property of $h(x,y)$, is its separability along the cross and along-track directions resulting in some simplifications of the analytical relationships governing the scanner operation. In practice, $h(x,y)$ is truncated at some point, usually $0.1 h(0,0)$, to keep the computation time down. The normalizing constant c_1 , provides a unity gain for this averaging operation (Appendix A).

An alternate PSF but not as yet operational aboard Landsat is the rectangular function defined by

$$h(x,y) = \begin{cases} 1/r_0^2 & |x|, |y| \leq r_0/2 \\ 1 & 0 \end{cases}$$

The definition of the IFOV adopted here for either a Gaussian or rectangular PSF is such that $\text{IFOV} = r_0$.

3.2.2 MSS Statistical Model and Spatial Correlation

As the input random processes undergo a linear transformation, so do their statistical properties. In order to investigate the various interactive relationships outlined previously, an understanding and knowledge of the signal flow through the scanner is essential.

Relating the statistics of the multispectral signal at the scanner output to the corresponding part at the input can be accomplished in various ways. It has been pointed out that a two dimensional convolution is equivalent to a matrix multiplication in which one matrix is block circulant [55]. Let F and G be the input and output matrices arranged in $P^2 \times 1$ column vectors. Then they are related by

$$G = HF \quad (3-3)$$

where PSF matrix H , has the following structure

$$H = \begin{bmatrix} H_0 & H_{P-1} & \cdots & H_1 \\ H_1 & H_0 & \cdots & H_2 \\ \vdots & \vdots & \ddots & \vdots \\ H_{P-1} & H_{P-2} & \cdots & H_0 \end{bmatrix}$$

Each element in H is itself a $P \times P$ matrix. For a particular case, a selected number of fields can be chosen and processed by (3-3) to produce the G matrix followed by the calculation of a pooled auto and cross spectral correlation matrix.

This method has the advantage of requiring no a priori spatial information yet its data dependent nature makes the results of any study limited to the particular data set used. The more general approach, providing possibly closed form expressions for the quantities desired, is the application of linear system theory techniques to the MSS. This, however, requires some a priori specification of data properties in an algebraic form, the main item being the spatial correlation model.

Agricultural crop planting, natural formations of terrain, water supplies, etc. all exhibit a certain homogeneity in their structure; therefore, it is expected that the reflected energy sensed by a scanner will show the same property in the form of a correlation between adjacent pixels of the final digital data set. Comparatively speaking, spectral classification has been much more widespread than

spatial classification, resulting in less than a full attention to the spatial properties of remotely sensed data. It has been suggested, however, that the experimentally observed correlation functions approximately follow a decaying exponential [56,57]. This assumption implies a Markov model for the spatial characteristics of the data. Let $\underline{R}_k$ be the spatial correlation matrix of the k th spectral band

$$\underline{R}_k = [r_{ij}] \quad i, j = 0, 1, \dots, n_0 - 1 \quad (3-5)$$

Under the two assumptions: (a), Markov correlation structure; and (b), separability along the cross-track and along-track directions, $\underline{R}_k$ can be specified as follows

$$\underline{R}_k = [r_{ij}] = \rho_{x_k}^i \rho_{y_k}^j \quad i, j = 0, 1, \dots, n_0 - 1 \quad (3-6)$$

where ρ_{x_k} and ρ_{y_k} are the adjacent pixel correlation coefficients along the respective directions given by

$$\begin{aligned} \rho_{x_k} &= e^{-a_{kk}} \\ \rho_{y_k} &= e^{-b_{kk}} \end{aligned} \quad (3-7)$$

Similarly, the spatial crosscorrelation matrix between two bands p and q is defined as

$$\underline{R}_{pq} = [r_{ij}] = \rho_{x_{pq}}^i \rho_{x_{pq}}^j \quad i, j = 0, 1, \dots, n_0 - 1 \quad (3-8)$$

where

$$\rho_{x_{pq}} = e^{-a_{pq}} \quad (3-9)$$

$$\rho_{y_{pq}} = e^{-b_{pq}}$$

In order to examine the validity of the Markov model and the separability property of the correlation functions, a sample aircraft MSS data set is selected and the estimate of the auto and crosscorrelation functions in two spectral bands, one in visible and one in near infrared, is obtained by a lagged-product sum method [58]. The separability characteristics can be checked by completing the entire correlation matrix, $\underline{R}$, using $[r_{ij}] = [r_i r_j]$ and comparing it to the experimentally observed quantity. Let E be the error matrix associated with this operation, then

$$E = |[r_{ij}] - [r_i r_j]| \quad (3-10)$$

The results are shown in Fig. 3-2 through Fig. 3-4 and Tables 3-1 through 3-3. Although the shape of the correlation curves themselves indicate an approximate exponential behavior, a quantitative weighted least-squares fit shows that this assumption is indeed valid. The differential between the correlation of the lines and columns of this data set stems from the fact that the analog signals are sampled in a way that generates unequal separation between the corresponding ground resolution

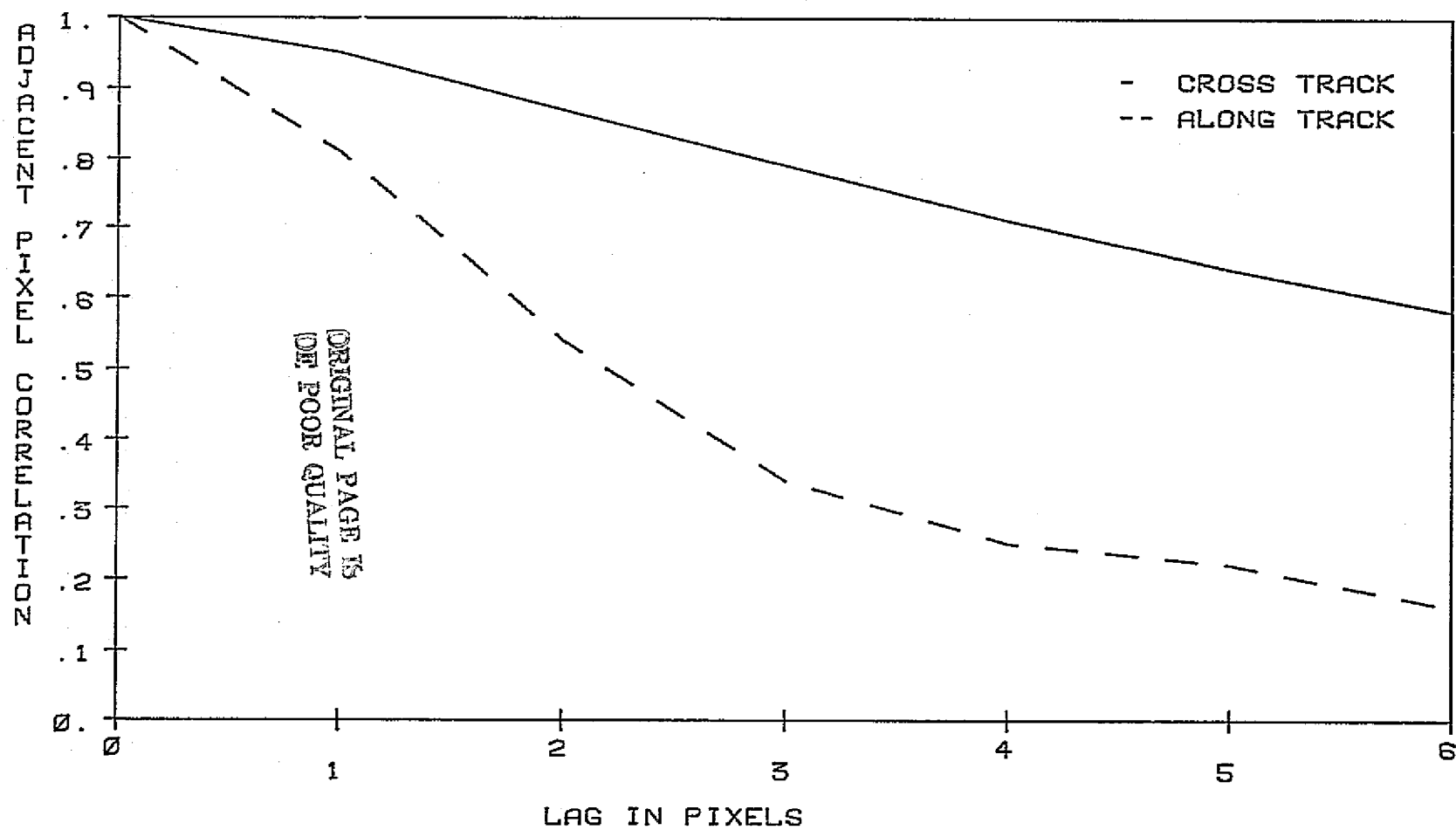


FIG. 3-2 AIRCRAFT MSS DATA AUTOCORRELATION FUNCTION. CHANNEL 2

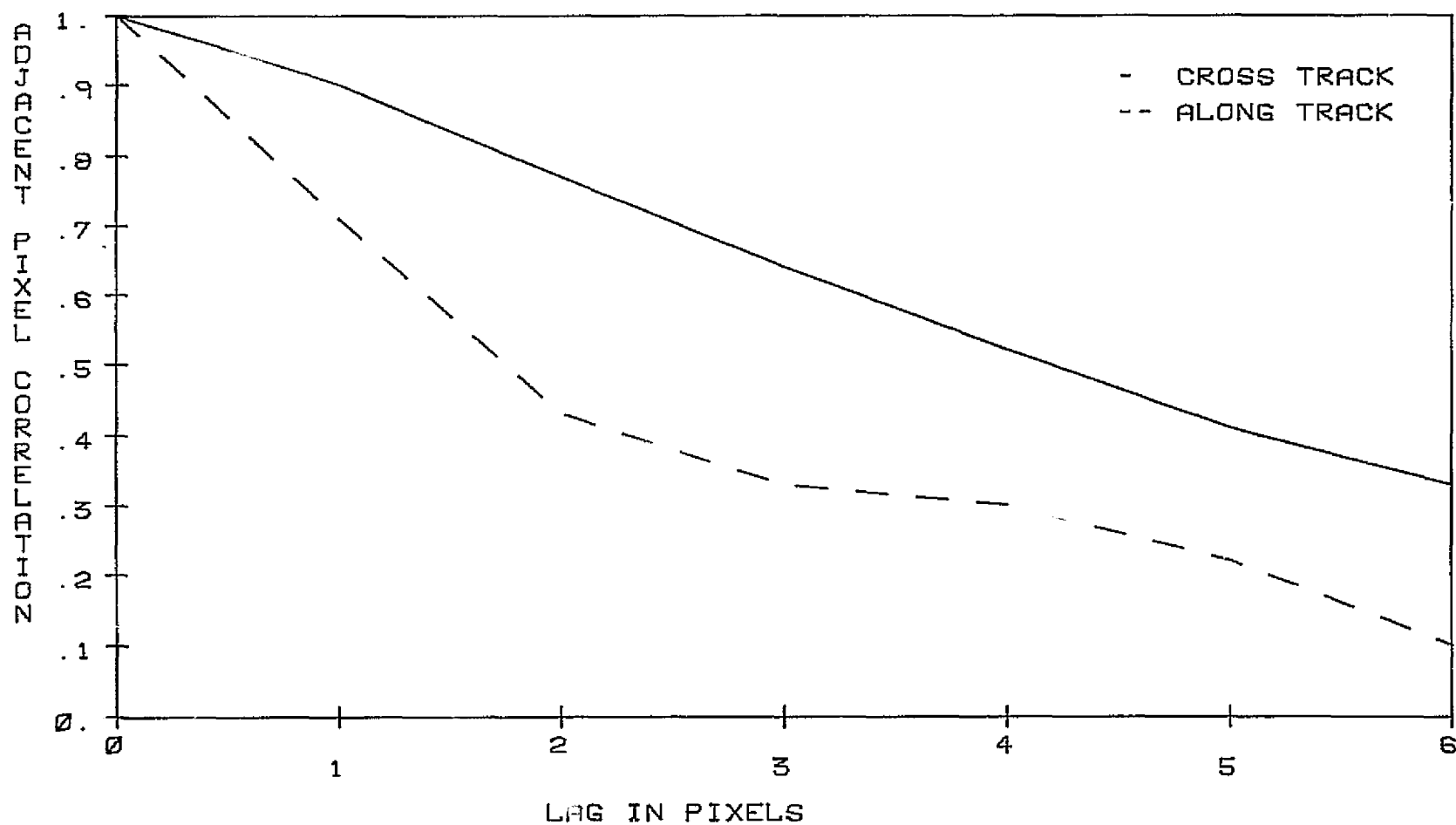


FIG. 3-3 AIRCRAFT MSS DATA AUTOCORRELATION FUNCTION. CHANNEL 9

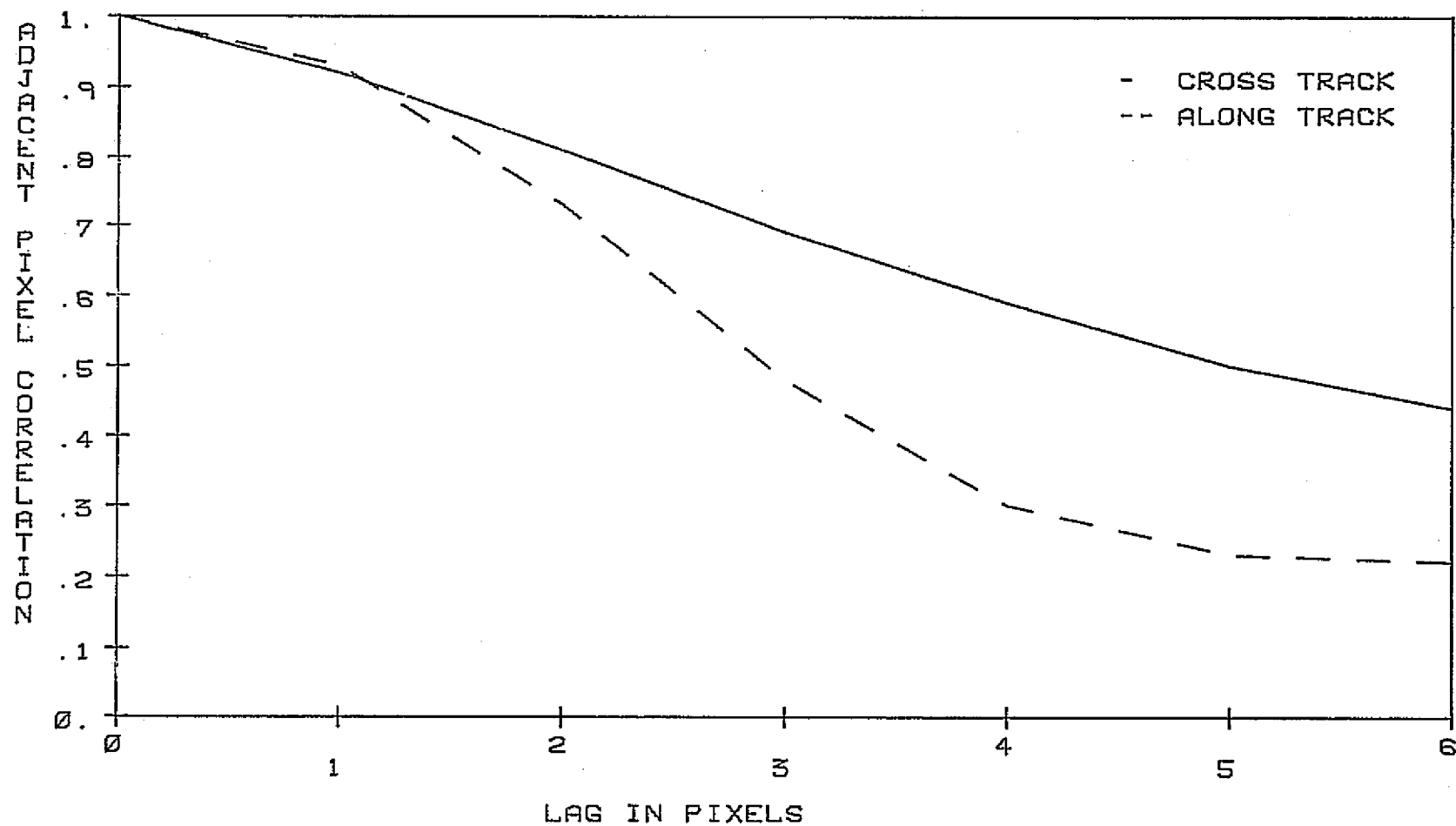


FIG. 3-4 AIRCRAFT MSS DATA CROSSCORRELATION FUNCTION. CHANNELS 2 AND 8

Table 3-1 Error Matrix for Correlation Function Approximation for Channel 2.

$$\underline{R}_2 = \begin{bmatrix} 1.00 & 0.95 & 0.87 & 0.79 & 0.71 & 0.64 & 0.58 \\ 0.81 & 0.79 & 0.74 & 0.67 & 0.60 & 0.53 & 0.48 \\ 0.54 & 0.55 & 0.52 & 0.48 & 0.42 & 0.31 & 0.32 \\ 0.34 & 0.36 & 0.35 & 0.33 & 0.29 & 0.25 & 0.21 \\ 0.25 & 0.27 & 0.28 & 0.27 & 0.25 & 0.21 & 0.18 \\ 0.22 & 0.24 & 0.25 & 0.25 & 0.23 & 0.21 & 0.18 \\ 0.16 & 0.19 & 0.20 & 0.21 & 0.20 & 0.19 & 0.16 \end{bmatrix}$$

$$\hat{\underline{R}}_2 = \begin{bmatrix} 1.00 & 0.95 & 0.87 & 0.79 & 0.71 & 0.64 & 0.58 \\ 0.81 & 0.77 & 0.70 & 0.63 & 0.57 & 0.51 & .47 \\ 0.54 & .51 & .47 & .42 & .38 & .34 & .31 \\ 0.34 & .32 & .3 & .27 & .24 & .21 & .2 \\ 0.75 & .23 & .21 & .19 & .17 & .16 & .14 \\ 0.22 & .2 & .19 & .17 & .15 & .14 & .12 \\ 0.16 & .25 & .13 & .12 & .11 & .1 & .09 \end{bmatrix}$$

$$\underline{E}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.6 & 5.4 & 5.9 & 5 & 3.7 & .2 \\ 0 & 7.2 & 9.6 & 12.5 & 9.5 & 8.1 & 3.1 \\ 0 & 11.1 & 14.2 & 18.1 & 17.2 & 16 & 4.7 \\ 0 & 14.8 & 25 & 29.6 & 32 & 23.8 & 22.2 \\ 0 & 16.6 & 24 & 32 & 34.7 & 33.3 & 33.3 \\ 0 & 24 & 35 & 42.8 & 45 & 47.3 & 43.7 \end{bmatrix}$$

Table 3-2 Error Matrix for Correlation Function Approximation for Channel 8.

R_{-8}	1.00	.9	.77	.64	.52	.41	.33
	.71	.68	.59	.48	.37	.28	.20
	.43	.43	.38	.3	.21	.12	.06
	.33	.35	.33	.27	.19	.12	.06
	.30	.32	.31	.28	.23	.17	.12
	.22	.75	.26	.24	.20	.15	.11
	.10	.13	.14	.13	.11	.08	.05

$\hat{R}_8$	1.00	.9	.77	.64	.52	.41	.33
	.71	.64	.55	.45	.37	.3	.11
	.43	.39	.33	.27	.22	.17	.14
	.33	.3	.25	.21	.17	.14	.11
	.30	.27	.23	.2	.15	.12	.09
	.22	.2	.17	.14	.11	.09	.07
	.1	.9	.07	.06	.05	.04	.03

E_8	0	0	0	0	0	0	0
	0	5.8	6.7	6.2	0	6.6	45
	0	9.3	13.1	10	4.5	30	57
	0	14.2	24.2	22.2	10.5	14.2	45.4
	0	15.6	25.8	28.5	34.7	29.4	25
	0	20	34.6	41.6	45	40	36.3
	0	30.7	50	53.8	54.5	50	40

ORIGINAL PAGE IS
OF POOR QUALITY

Table 3-3 Error Matrix for Cross Correlation Function Approximation
Between Channels 2 and 8.

$$R_{28} = \begin{bmatrix} 1.00 & .92 & .81 & .69 & .59 & .50 & .44 \\ .93 & .88 & .78 & .67 & .56 & .48 & .41 \\ .73 & .71 & .64 & .54 & .44 & .36 & .3 \\ .48 & .47 & .43 & .36 & .28 & .21 & .16 \\ .30 & .31 & .29 & .24 & .18 & .12 & .08 \\ .23 & .25 & .24 & .21 & .16 & .12 & .08 \\ .22 & .24 & .24 & .22 & .19 & .15 & .12 \end{bmatrix}$$

$$\hat{R}_{28} = \begin{bmatrix} 1.00 & .92 & .81 & .69 & .59 & .50 & .44 \\ .93 & .86 & .75 & .64 & .55 & .46 & .4 \\ .73 & .67 & .6 & .5 & .43 & .36 & .32 \\ .48 & .44 & .38 & .33 & .28 & .24 & .21 \\ .30 & .27 & .24 & .2 & .17 & .15 & .13 \\ .23 & .21 & .18 & .16 & .13 & .11 & .1 \\ .22 & .2 & .18 & .15 & .13 & .11 & .1 \end{bmatrix}$$

$$E_{28} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.2 & 3.8 & 4.5 & 1.8 & 4.16 & 2.43 \\ 0 & 5.6 & 6.2 & 7.4 & 2.3 & 0 & 6.75 \\ 0 & 6.4 & 11.6 & 8.3 & 0 & 17.5 & 23.8 \\ 0 & 13.0 & 17.2 & 16.6 & 5.5 & 20 & 38.4 \\ 0 & 16 & 25 & 23.8 & 18.7 & 8.3 & 20 \\ 0 & 16.6 & 25 & 31.8 & 31.6 & 26.6 & 16.6 \end{bmatrix}$$

elements along the scan swath and vehicle down track motion. The unusually high cross-track pixel-to-pixel correlation is attributed to the use of very high resolution aircraft data. For satellite imagery a $\rho_{x_k} = 0.8$ is a more common value.

The separability property of the correlation matrices appears to be a reasonable assumption according to the correlation error matrices. As will be shown later, this is not a feature peculiar to this data set but is observed throughout most of the multispectral data bases. The main property exhibited by E, is that the separability assumption becomes progressively invalid for higher lag values. This, however, is not particularly detrimental to the correlation model proposed here due to the fact that although the absolute error term expressed in percentage can be relatively high, the normalized values of the correlation function in the range of concern are themselves quite small, and thus, carry little weight in influencing the final results.

With the correlation model well defined, the output spectral covariance matrix can be specified. Let $R_{g_i g_j}$ and $\sum_g$ be the output spatial correlation matrix between spectral bands i and j and output covariance matrix, respectively, then

$$\sum_g(i,j) = [R_{g_i g_j}(0,0)] \quad i,j=1,2,\dots,N \quad (3-11)$$

Note that when considered over the ensemble of all the bands, matrix R_g is an $(n_o \times N)(n_o \times N)$ partitioned matrix, given by

$$\underline{R}_g = \begin{bmatrix} [R_{g_1 g_1}] & [R_{g_1 g_2}] & \dots & [R_{g_1 g_N}] \\ [R_{g_2 g_1}] & [R_{g_2 g_2}] & \dots & [R_{g_2 g_N}] \\ \vdots & \vdots & \ddots & \vdots \\ [R_{g_N g_1}] & \dots & \dots & [R_{g_N g_N}] \end{bmatrix} \quad (3-12)$$

where $[R_{i,j}]$ is the $n_o \times n_o$ spatial correlation matrix. $\sum_g$ however, is only a function of zero lag elements of R_g , $R_{g_i g_j}(0,0)$. Therefore, only $N \times N$ out of $(n_o \times N)(n_o \times N)$ entries of $\underline{R}_g$ need be calculated. It is clear that the spectral correlation matrix is a small subset of spatial correlation matrices whose elements have the following locations.

$$\sum_g(i,j) = R_g(i-1)n_o, (j-1)n_o) \quad i,j=1,2, \dots, N \quad (3-13)$$

The analytical relationship between the input and output correlation matrices of an N-band MSS is investigated in Appendix A. Specific results are obtained for a Markov-correlated data set, a Gaussian and a rectangular shaped scanner IPOV. The main result obtained there is a scanner characteristic function $W_s(\tau, \eta, a, b)$ given by

$$W_s(\tau, \eta, a_{ii}, b_{ii}) = \left[e^{\frac{a_{ii}^2 r_o^2}{2} - a_{ii} \tau Q(a_{ii} r_o - \frac{\tau}{r_o})} + e^{\frac{a_{ii}^2 r_o^2}{2} + a_{ii} \tau Q(a_{ii} r_o + \frac{\tau}{r_o})} \right] \times$$

$$\left[e^{\frac{b_{ii}^2 r_o^2}{2} - b_{ii} \eta Q(b_{ii} r_o - \frac{\eta}{r_o})} + e^{\frac{b_{ii}^2 r_o^2}{2} + b_{ii} \eta Q(b_{ii} r_o + \frac{\eta}{r_o})} \right]$$

(3-14)

where a_{ii} and b_{ii} are the parameters of input spatial correlation function determining the adjacent pixel correlation in band i , r_o is the scanner PSF characteristic length and Q is as defined earlier.

W_s plays a central role in the spatial modeling of a multispectral scanner. It is a function by which all channel variances and band-to-band correlation coefficients are weighted to produce the corresponding output quantity. Specifically,

$$\sigma_{g_i}^2 = W_s(0, 0, a_{ii}, b_{ii}) \sigma_{f_i}^2 \quad i=1, \dots, N$$

$$s_{g_i g_j} = \frac{W_s(0, 0, a_{ij}, b_{ij})}{W_s^{\frac{1}{2}}(0, 0, a_{ii}, b_{ii}) \times W_s^{\frac{1}{2}}(0, 0, a_{jj}, b_{jj})} s_{f_i f_j} \quad \begin{matrix} i, j=1, \dots, N \\ i \neq j \end{matrix}$$

$$W_s^{\frac{1}{2}}(0, 0, a_{jj}, b_{jj}) \quad (3-15)$$

ORIGINAL PAGE IS
OF POOR QUALITY

where $s_{f_i f_j}$, $s_{g_i g_j}$, (a_{ii}, b_{ii}) and (a_{ij}, b_{ij}) are the input crosscorrelation coefficient between bands i and j . The corresponding output quantity, the parameters of the band i autocorrelation function and the parameters of bands i and j crosscorrelation function, respectively.

Evaluating $W_s(\tau, \eta, a, b)$ for all values of τ and η can complete the entire output spatial matrix R_g . The Bayes classifier, however, is not a spatial classifier but, rather, is a spectral one and, as a result, the knowledge of an $N \times N$ spectral covariance matrix is sufficient for classification purposes. As it was envisioned at the beginning, developing a parametric model provides a significant flexibility in the system analysis. For example, W_s can selectively supply any entry of the output spatial matrix desired. Here, $W_s(\tau, \eta, a, b) \Big|_{\tau=\eta=0}$ can complete the output spectral covariance matrix

$$W_s(0, 0, a, b) = 4e^{\frac{(a^2+b^2)}{2} r_o^2} Q(ar_o) Q(br_o) \quad (3-16)$$

For example, when the input random process is a two spectral band data set, the output spectral correlation matrix, S_g is given in terms of S_f as follows:

$$\begin{aligned}
 \underline{s}_f &= \begin{bmatrix} 1 & s_{f_1 f_2} \\ & 1 \end{bmatrix} \\
 \underline{s}_g &= \begin{bmatrix} 1 & \frac{W_s(0,0,a_{12},b_{12})}{W_s^{\frac{1}{2}}(0,0,a_{11},b_{11})W_s^{\frac{1}{2}}(0,0,a_{22},b_{22})} s_{f_1 f_2} \\ & 1 \end{bmatrix}
 \end{aligned}$$

(3-17)

It is clear that, depending on the particular value of W_s , the output correlation matrices, and hence, classification accuracies will be modified. The variations of W_s as a function of scene correlation and scanner spatial parameters can be very illuminating. For a Gaussian scanner PSF, W_s is plotted vs. the sample-to-sample correlation for a fixed line-to-line correlation. The IFOV is used as a running parameter, Fig. 3-5 through 3-12. The adjacent sample correlation coefficient ranges from a near white noise 0.1 to total correlation of 1 (constant signal amplitude). The adjacent line correlation coefficient extends from 0.65 to 1. Similar plots are shown for a rectangular PSF, Fig. 3-13 through 3-15. Examination of these results reveals several important features: (a), Since $0 \leq W_s \leq 1$, the output channel variances are always smaller than the corresponding input quantity. This is an expected result due to

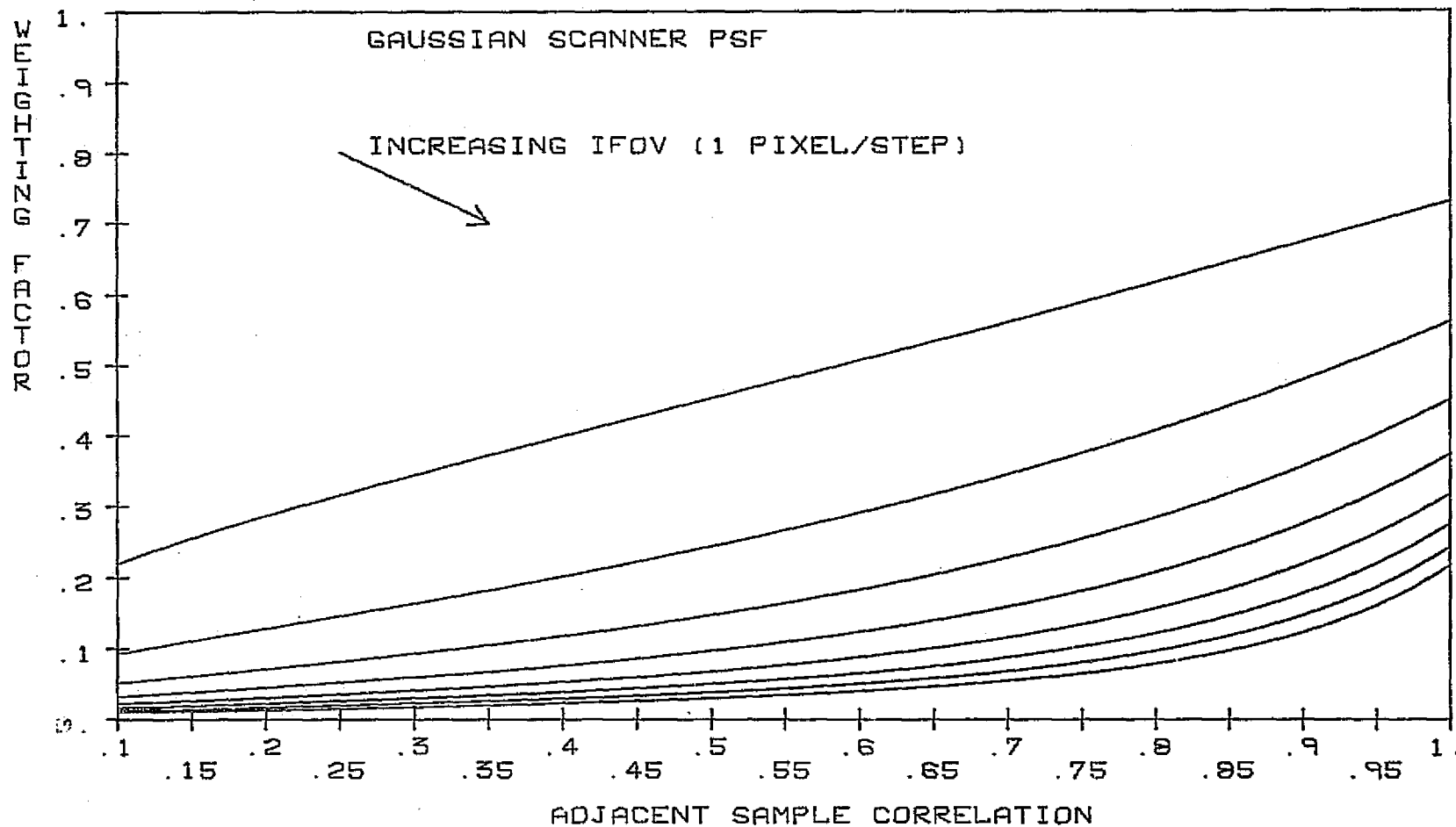


FIG. 3-5 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION = .65

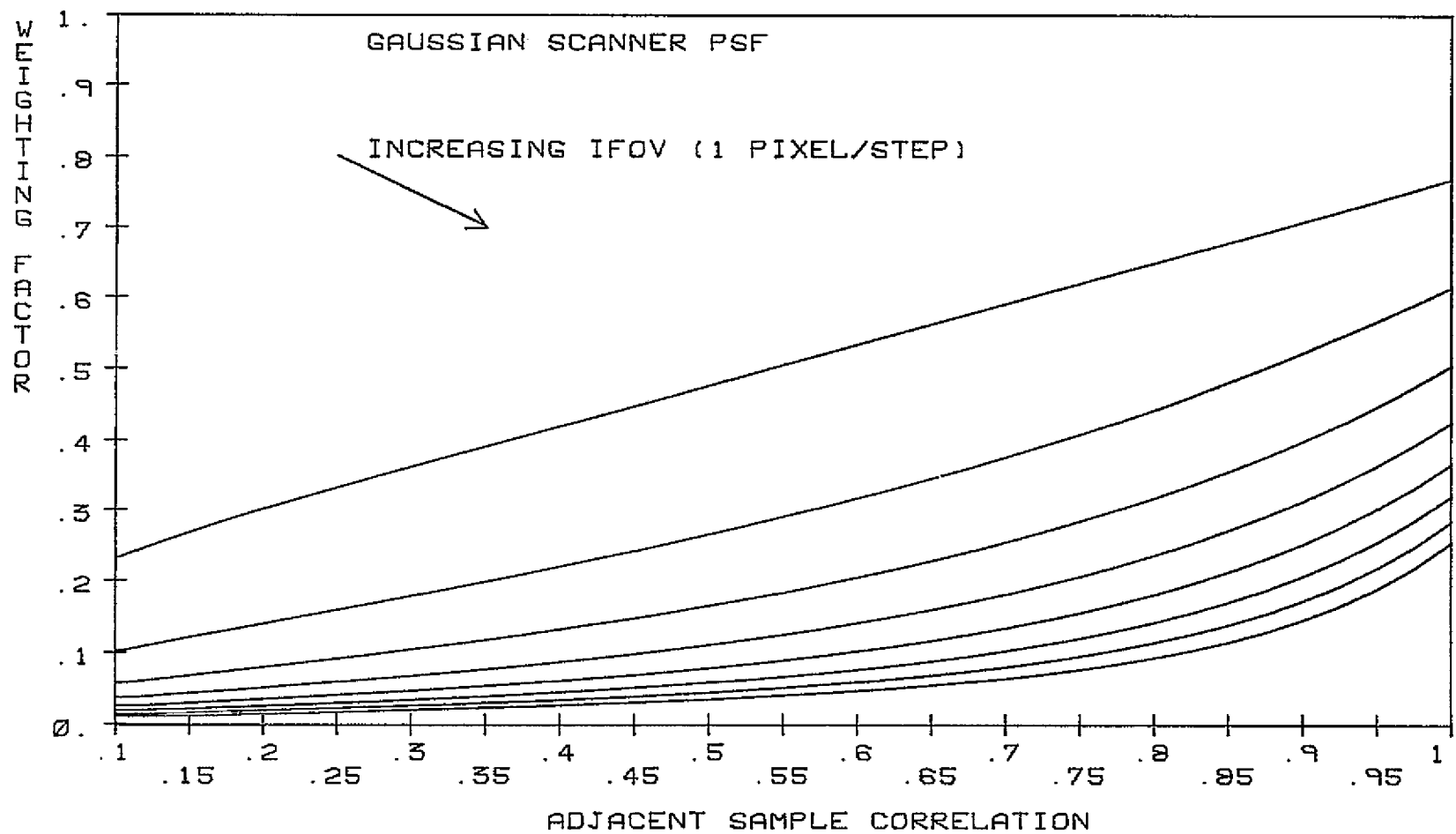


FIG. 3-6 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION = .7

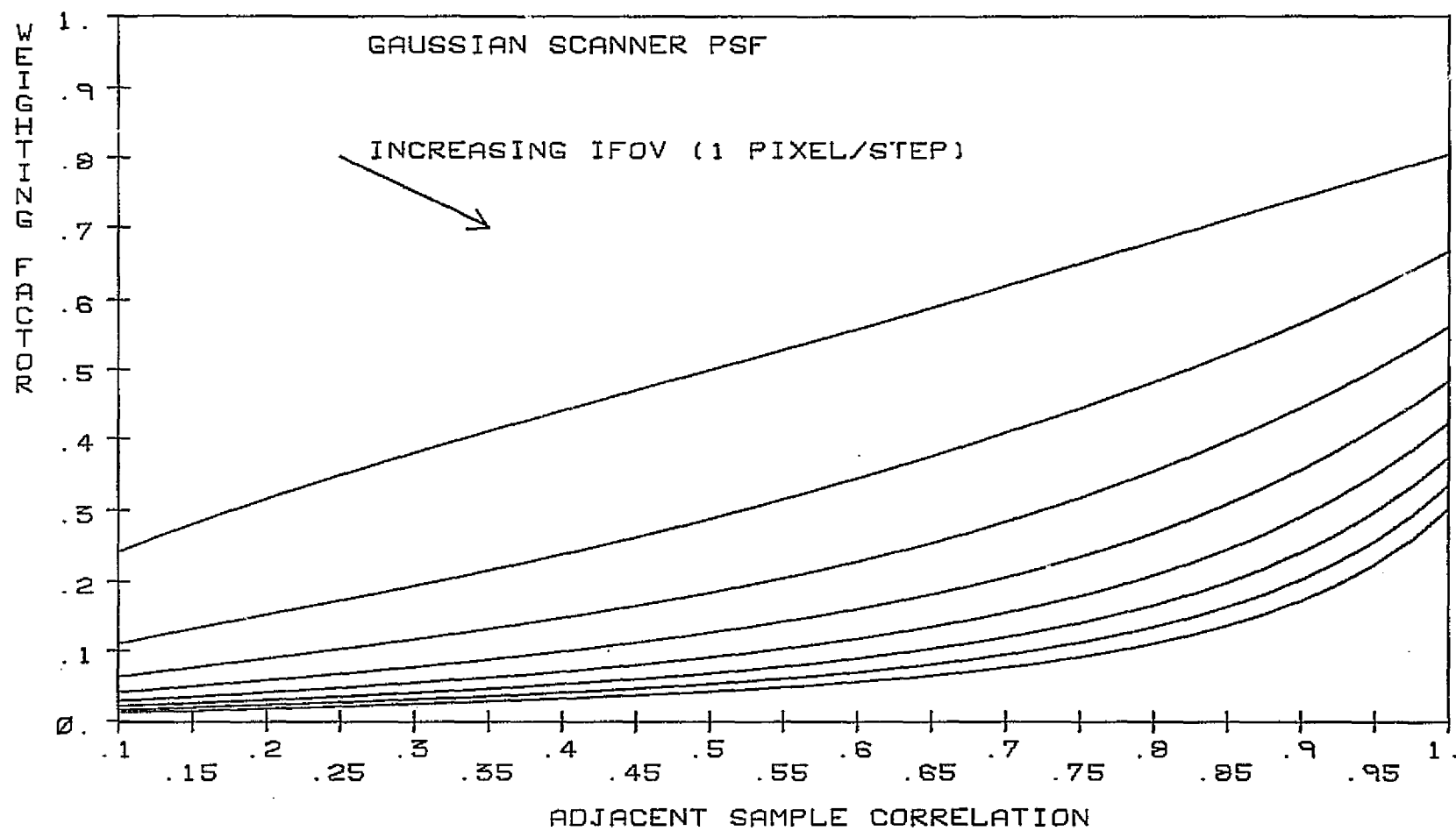


FIG. 3-7 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION = .75

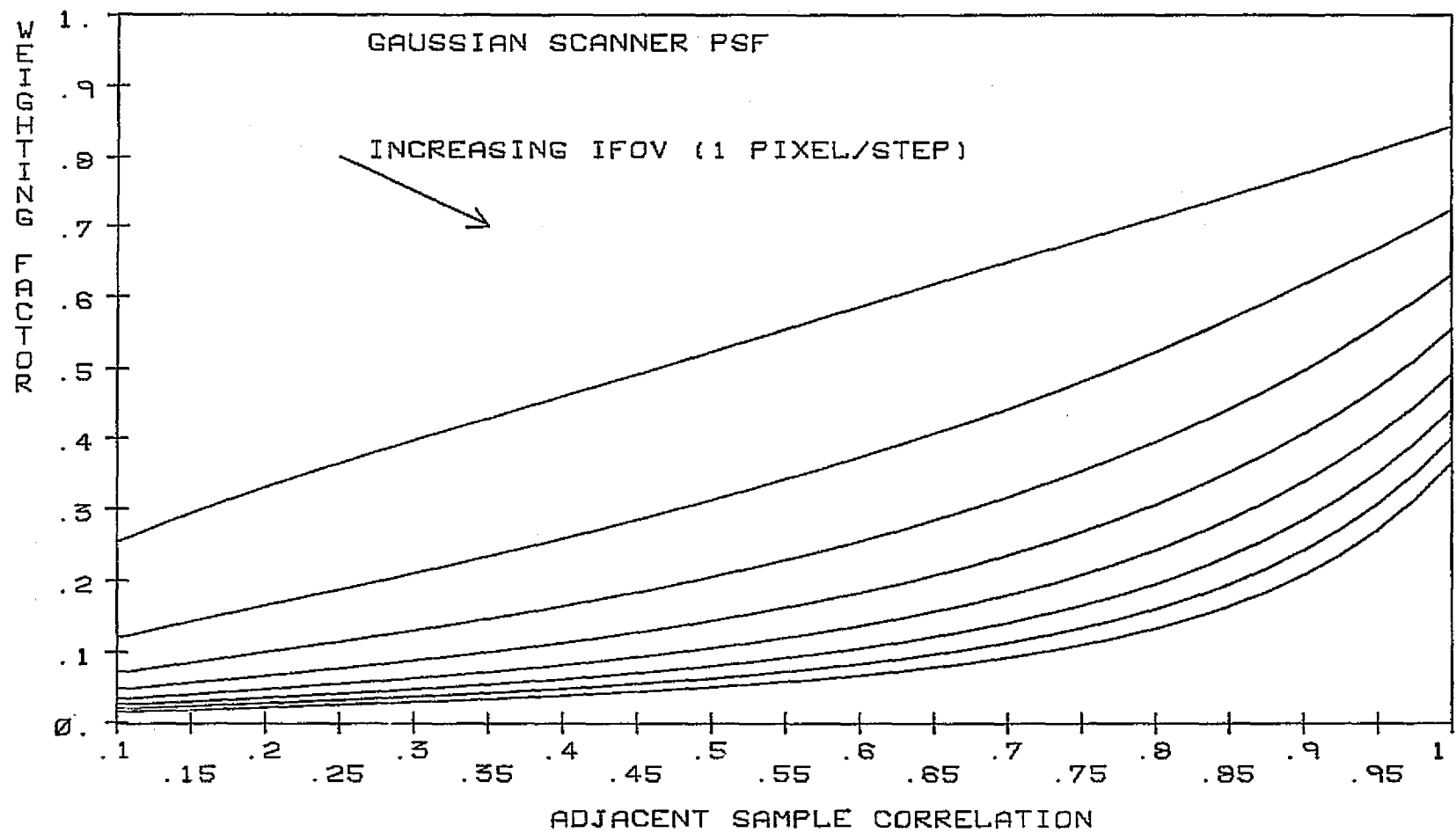


FIG. 3-8 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION = .8

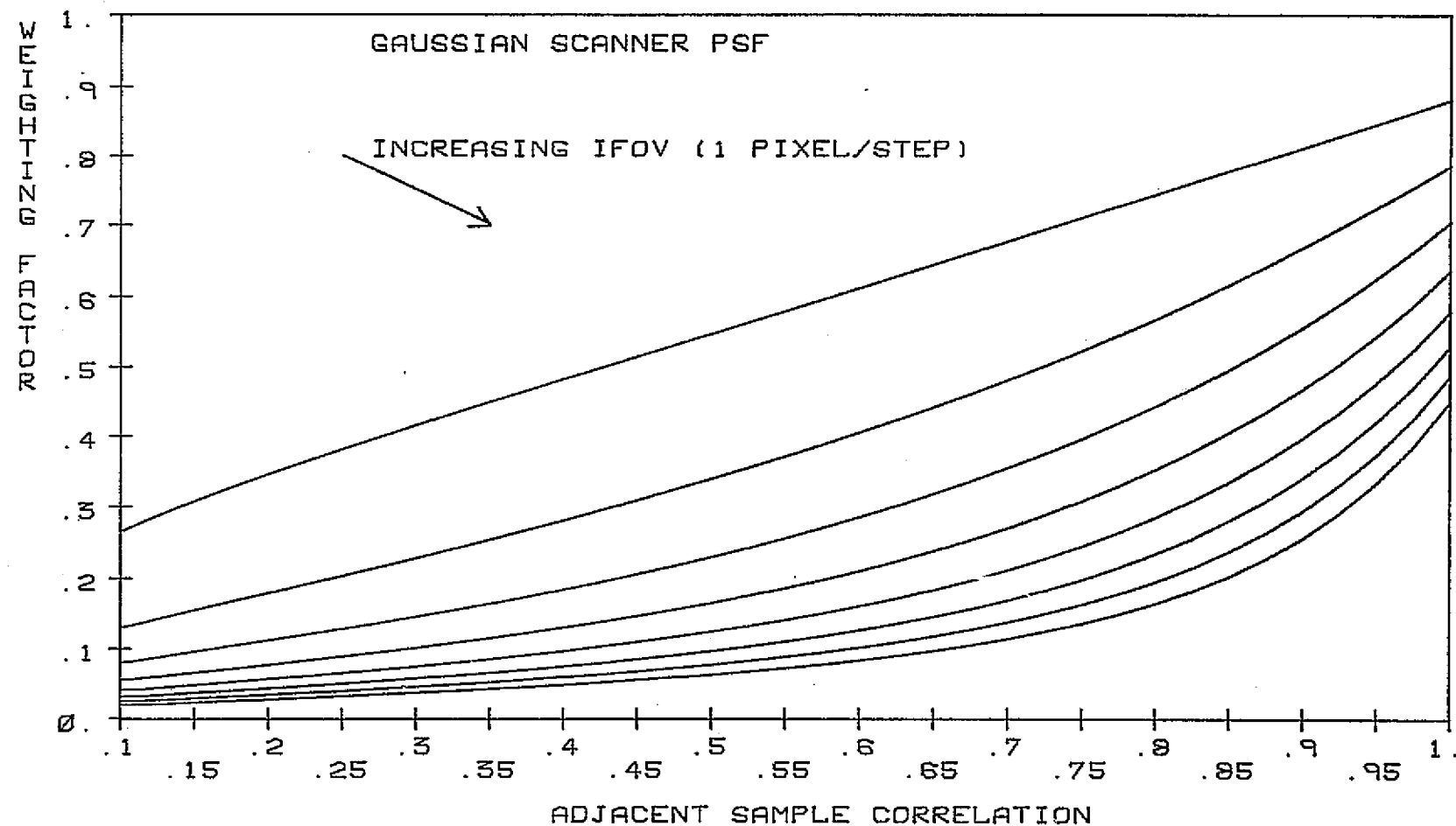


FIG. 3-9 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION = .85

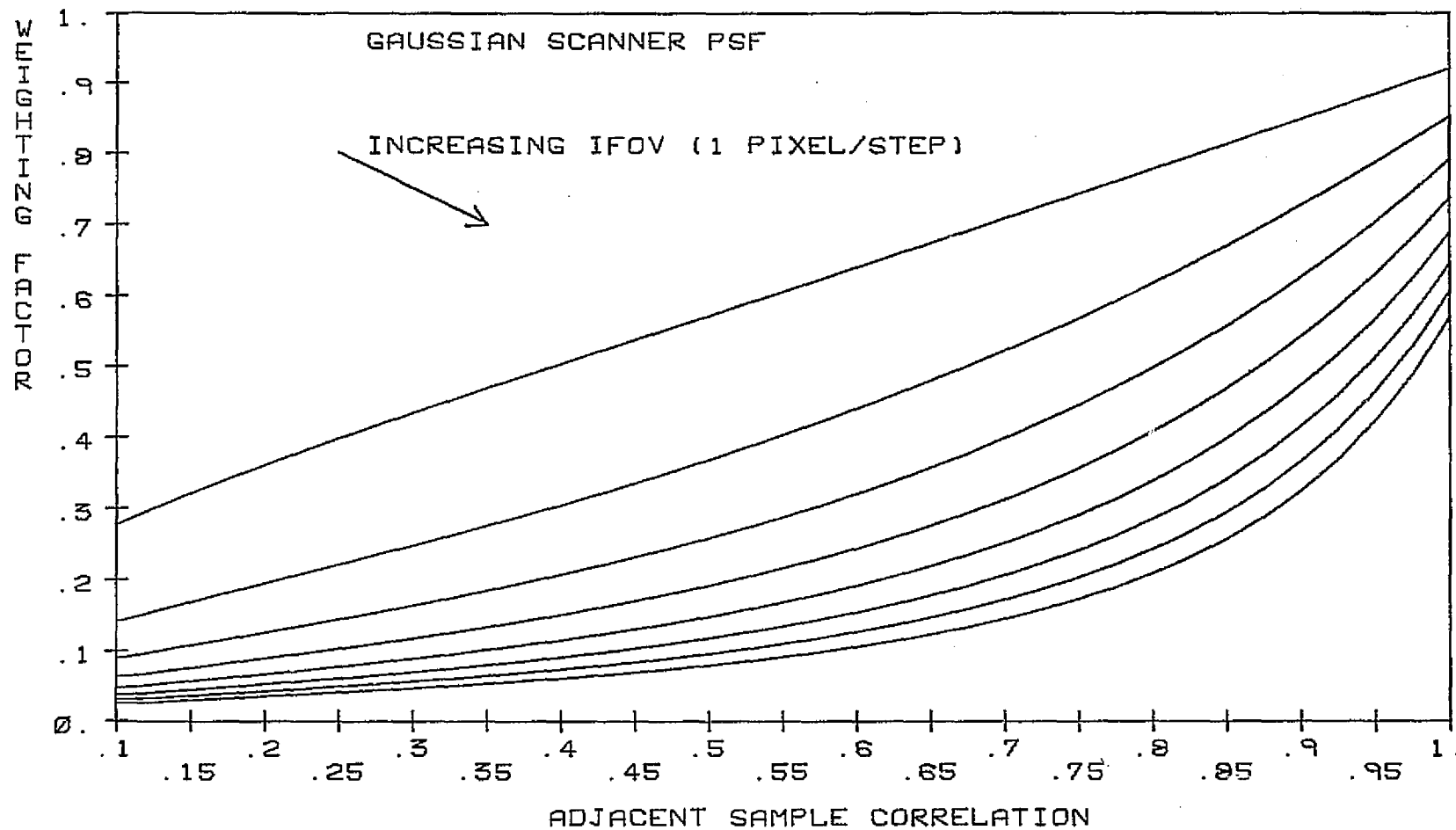


FIG. 3-10 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION = .9

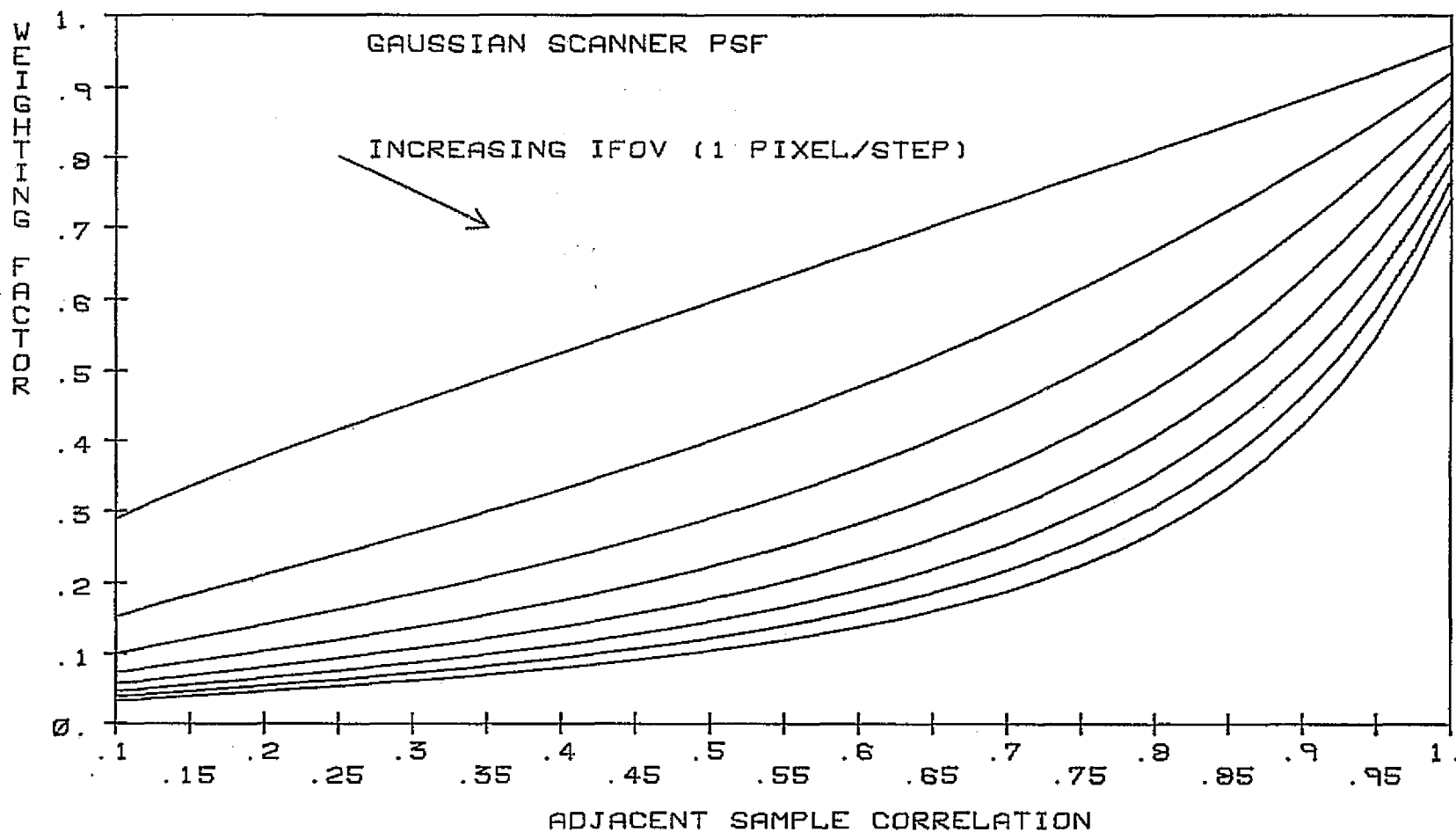


FIG. 3-11 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION = .95

ORIGINAL PAGE IS
OF POOR QUALITY

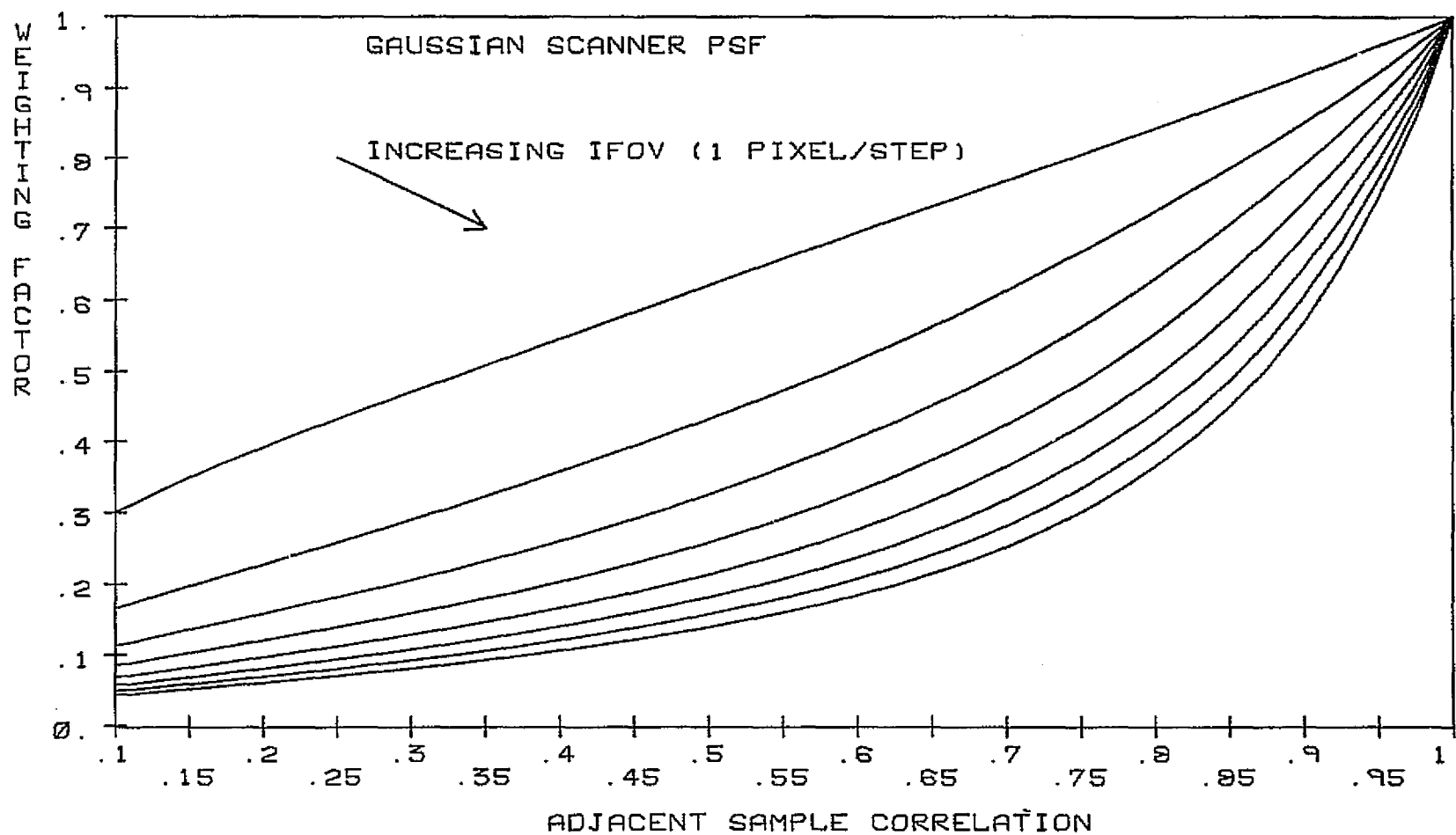


FIG. 3-12 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION = 1.

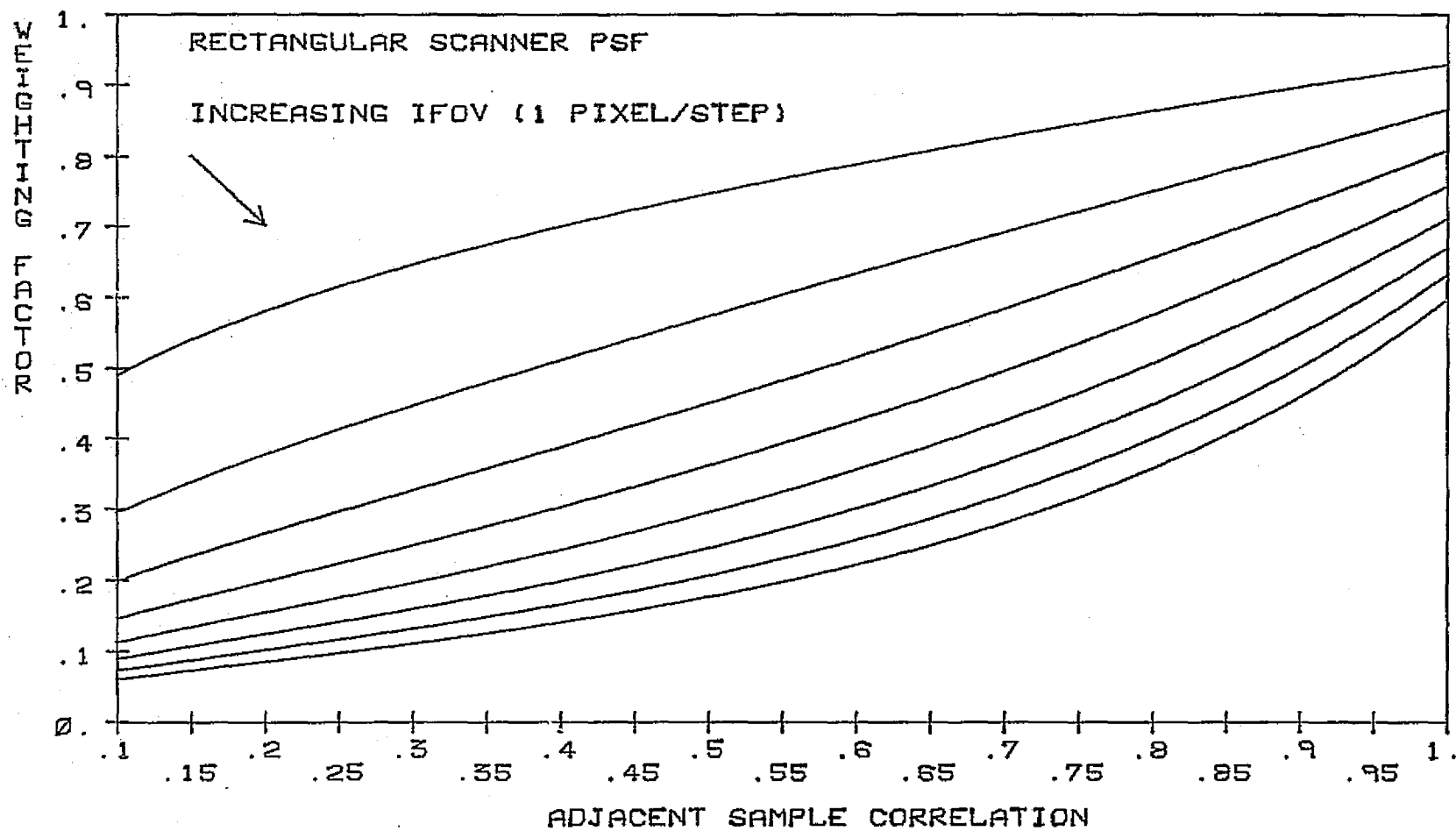


FIG. 3-13 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION=.8

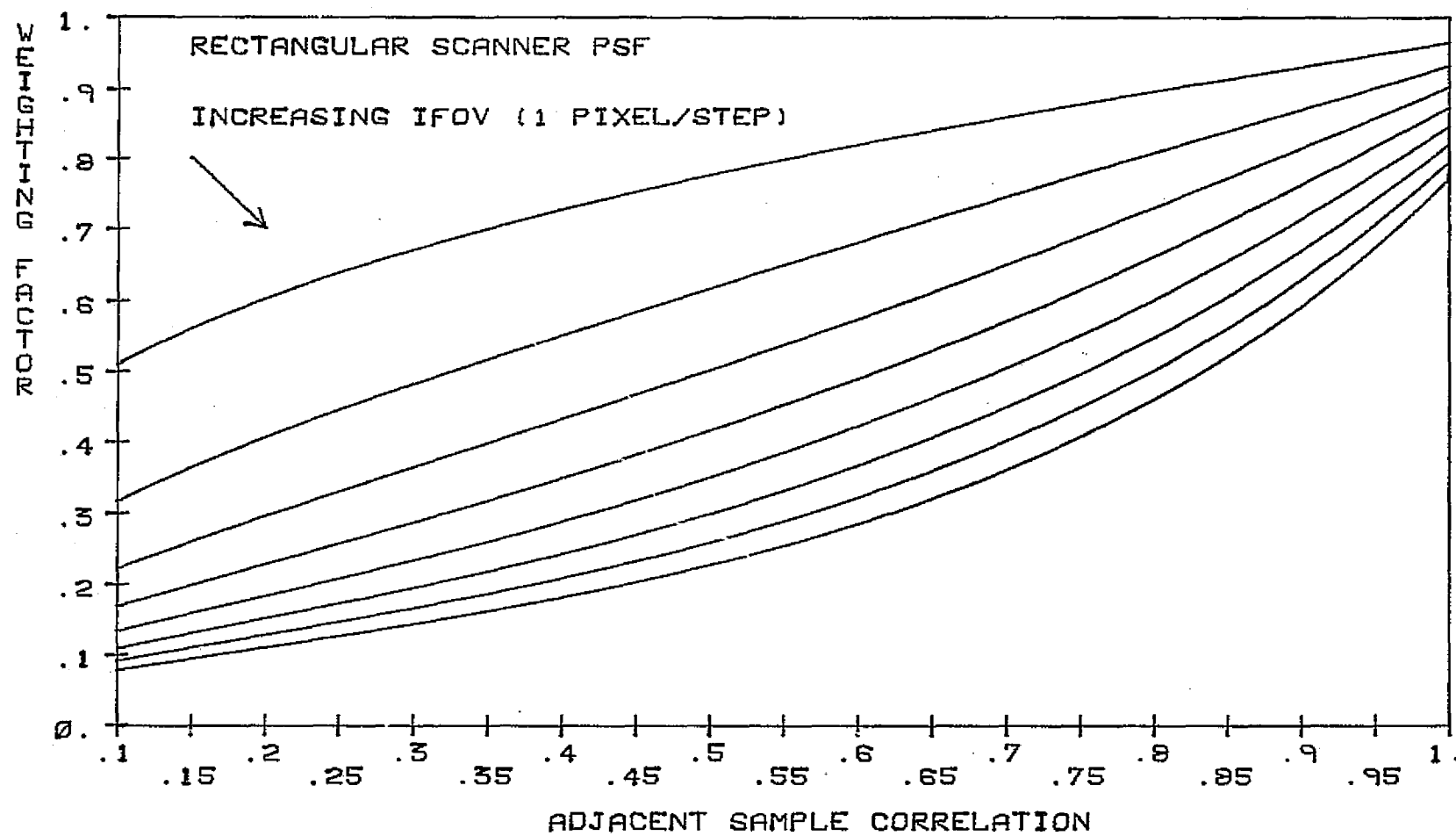


FIG. 3-14 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION=.9

ORIGINAL PAGE IS
OF POOR QUALITY

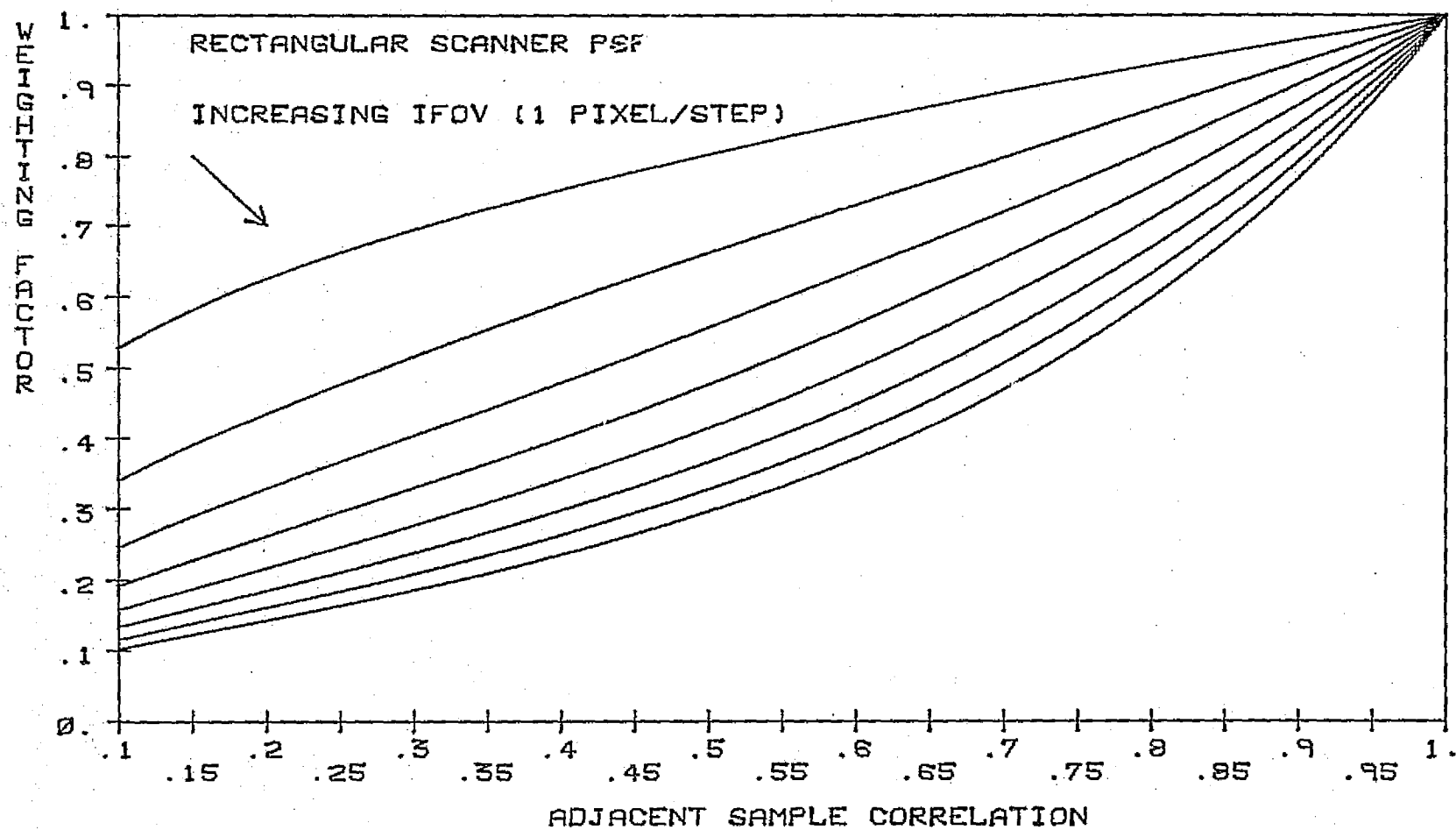


FIG. 3-15 SCANNER CHARACTERISTIC FUNCTION VS. SCENE CORRELATION
ADJACENT LINE CORRELATION=1.

the averaging operation of the scanner, (b), for a fixed sample-to-sample correlation the spectral band variances at the output increase with decreasing IFOV with an accompanying degradation in classification accuracy, (c), for a fixed IFOV, the channel variances increase with decreasing scene correlation. These observations apply to any one of the cases with a fixed inter-line correlation. Consider two cases in which IFOV and sample-to-sample correlation are fixed, then a higher adjacent line correlation produces an increase in the output band variance.

The variations of the spectral correlation coefficients between bands are similarly determined. From (3-17), depending on the parameters of the correlation model, the ratio of two characteristic functions can potentially either increase or decrease the spectral band correlation.

3.3 Noise in Multispectral Scanner System Modeling

Random noise is the ultimate limiting factor in a data transmission and processing system. Although the performance of remote sensing systems is affected by many other parameters, additional noise entering the system at various stages can have a significant impact on the final analysis of the data. Hence, no model would be complete without the identification of the noise sources and determination of their contribution to the system performance degradation.

There are two broad categories of noise generating sources: external and internal. External noise is primarily caused by the atmosphere in the form of molecular absorption and scattering. In the case of the MSS in Landsat there are two major absorption bands at a wavelength of about $0.68\text{ }\mu\text{m}$ due to the presence of oxygen and water vapor which result in an attenuation of up to 10% or more for a vertical path from the surface of the earth to the platform. Scattering is the major cause of attenuation of the reflected energy. It has been experimentally observed that a combined Rayleigh and Mie scattering can cause up to 40% transmission loss through the atmosphere at $0.4\text{ }\mu\text{m}$ with a decreasing effect at higher wavelengths [59]. A designer has little influence over these natural phenomena and can only select appropriate windows in the atmospheric transmission spectrum to minimize absorption and scattering. In view of this situation consideration of external noise sources will not be pursued further.

3.3.1 System Noise

The noise generated within the scanner subsystem is primarily of two types: (a) noise introduced by the sensors in the detection stage of the incoming radiation; and (b) the quantization noise developed in the A/D conversion process prior to transmission to the ground stations.

Detectors are the most basic and crucial elements in a scanner system. Initially, thermal detectors, in which the impinging radiation heats a sensitive element and a temperature-dependent property is monitored, were in widespread use. The advent of high speed scanning mechanisms, requiring extremely short dwell time on a ground resolution element, required detectors with much higher sensitivity than thermal detectors. Photodetectors, where the photon energy in the incident radiation produces free charge carriers, are now primarily used in visible and infrared detection stages and provide time constants of the order of nanoseconds. Their disadvantage, compared to thermal detectors, is their limited spectral response and in most cases they require cooling. The currently operational Landsat-2 employs photomultipliers for the bands 0.5-0.6 μm , 0.6-0.7 μm , 0.7-0.8 μm , and silicon photodiodes for the range 0.8-1.1 μm . Landsat-C will carry a thermal band, 10.4-12.6 μm using two mercury-cadmium-telluride detectors [60].

The noise generated by a detector is a combination of photon and photomission noise. Let $\epsilon < 1$ be the photocathode efficiency of a photomultiplier with a gain G^n , the sampling time T , the charge on an electron $q_e = 1.6 \times 10^{-19}$ coulomb and the signal current out of the detector, I_s . The signal-to-noise power ratio at the output is given by [61].

$$\text{SNR} = I_s (G-1) T / q_e (1+\epsilon) G^{n+1} \quad (3-18)$$

The sampling time per detector, T , for the Landsat MSS is about $0.4 \mu\text{s}$. Assuming some typical values for other parameters:

$$I_s = 1 \text{ mA}$$

$$G = 3$$

$$n = 10$$

the SNR at the detectors output is approximately

$$\text{SNR} \approx 42 \text{ dB} \quad (3-19)$$

The next noise source is the A/D conversion process where analog signals are sampled and quantized to 2^B levels, each B bits long. The performance of the quantizer can be, evaluated in two ways. It is clear that the signal presented to the digitizer is already corrupted by detector noise, so the signal plus noise is actually being quantized and assigned to one of the 2^B levels. Therefore, the presence of noise makes this assignment subject to a finite probability of error thus affecting the performance measure. The second method simply involves the specification of noise power introduced by a uniform quantizer and is given by [62]

$$\sigma_n^2 = \Delta^2/12 \quad (3-20)$$

where Δ is the quantization step size. Defining a balanced system in which the detector and quantization noise are

equal, the combined SNR is therefore

$$\text{SNR} \approx 39 \text{ dB} \quad (3-21)$$

The data generated by Landsat is quantized to one of 64 levels (6 bit per pixel) with $\Delta=1$. In terms of the first performance measure, the assumption of equality of quantization and detector noise contribution to total system noise, implies that at the quantizer input

$$\psi = \frac{\Delta}{\sigma_n} = \sqrt{12} \quad (3-22)$$

where ψ is the ratio of step size to rms noise. For this particular value of ψ , the probability of the 6th bit being incorrectly assigned is 0.12 and essentially zero for 6th and/or the 5th bit [63,61].

Random noise in the context of multispectral remotely sensed data takes on a particular role. In the more classical applications of pattern recognition such as an M-ary communication channel employing one of M equally likely and known signals, noise is identified as the primary limiting factor in detecting the transmitted message with zero probability of error. The distinction emerges at this point that multispectral data is itself a realization of a stochastic process and as such, there is an inherent finite probability of error, regardless of noise, associated with the testing of hypothesis. In the analysis of the data, the noise and signal statistics will be merged and represent

ORIGINAL PAGE IS
OF POOR QUALITY

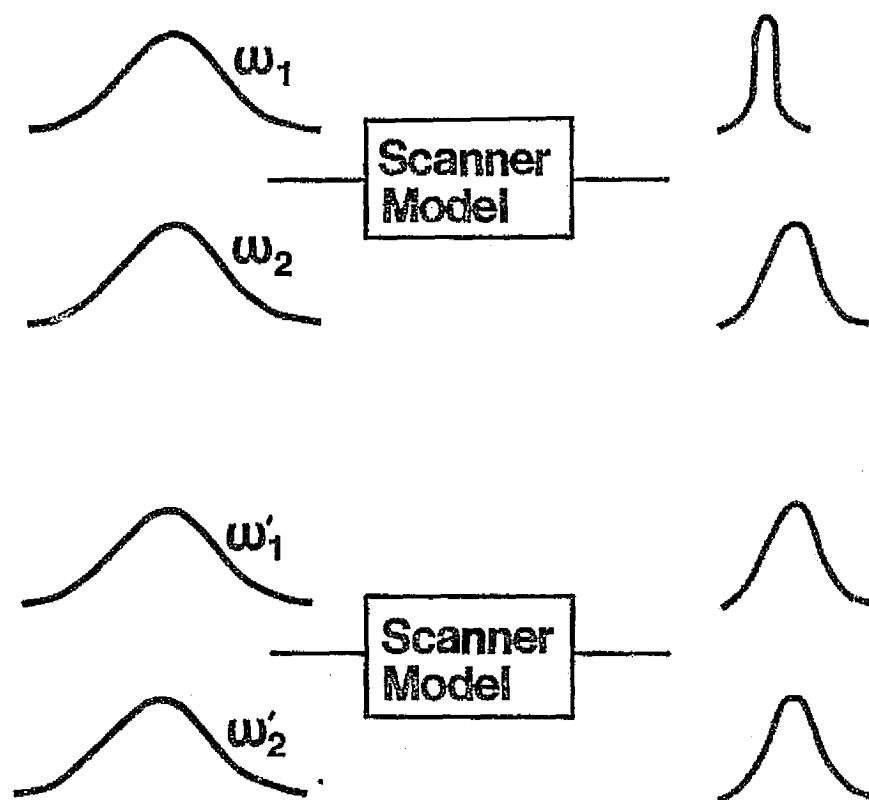


Fig. 3-16 Effect of Input Noise on Scanner Output Class Separability.

the population statistics and if the additive noise has Gaussian properties, the populations will still be normally distributed. In fact, it is plausible for some other hypothetical class to have identical statistics without noise as a part of its own properties. Therefore, a 'noisy' class can potentially be as separable as another 'noise-free' population.

The additive random noise has two major impacts on the statistics of multispectral data. The obvious one is the broadening of the distributions, resulting in more inter-class overlap hence a higher error rate. The second effect is on the data spatial correlation where the adjacent pixel correlation decreases with increasing noise power. Consider two univariate populations, Fig. 3-16, with equal variances where the first class is corrupted by random additive Gaussian noise. After transmission through the scanner, according to the properties of W_s , ω_1 emerges with a smaller variance than ω_2 with corresponding classification accuracies, $P_{c|\omega_1}$ and $P_{c|\omega_2}$. Consider two other populations, ω'_1 and ω'_2 with identical variances such that

$$\text{Var}\{\omega'_1\} = \text{Var}\{\omega'_2\} = \text{Var}\{\omega_1\} = \text{Var}\{\omega_2\} \quad (3-23)$$

where neither ω'_1 or ω'_2 are affected by random noise. Therefore,

ORIGINAL PAGE IS
OF POOR QUALITY

$$\rho_{x|\omega_1'} > \rho_{x|\omega_1} \quad (3.24)$$

and similarly for $\rho_{y|\omega_1}$. At the scanner output, the classification accuracies are $P_{c|\omega_1'}$ and $P_{c|\omega_2'}$. From the spatial correlation properties of ω_1' and ω_2' expressed in (3-24) and W_s it is clear that

$$P_{c|\omega_1'} > P_{c|\omega_1} \quad (3-25)$$

Does this mean the noisier the data the better? In terms of intrinsic classifiability of a population maybe but the question is *what* is being classified. Random noise can alter a statistics to the point that it will no longer represent the specific class under consideration and in fact in a multipopulation environment, the modified statistics could approach those of another existing population and thus increase the overall error rate not to mention the esthetic degradation of the image caused by it.

Another topic to be considered and defined is the term signal-to-noise ratio. It is frequently desirable to examine the performance of a system in a variable noise content environment. When the subject is the actual data, it should be noted that one is already dealing with a

noisy signal and, therefore, any additive noise will be in addition to the existing quantity. Let R be the noisy data, S the signal and N the noise, then

$$R(x,y) = S(x,y) + N(x,y) \quad (3-26)$$

The artificial noise N' is added to R to produce R'

$$R'(x,y) = R(x,y) + N'(x,y) \quad (3-27)$$

The $(SNR)' = \sigma_R^2 / \sigma_{N'}^2$ and $SNR = \sigma_S^2 / \sigma_N^2$ are related by

$$(SNR)' = \frac{\sigma_N^2}{\sigma_{N'}^2} + SNR \quad (3-28)$$

If the noise content of the data is considerably smaller than the added noise, then $(SNR)' \approx SNR$.

The way to determine the noise power to be added to the multispectral data for simulation purposes is open to discussion. Consider a frame of data, $R(x,y)$, containing M populations. A particular SNR can be specified and from that the noise variance derived. The signal variance, however, is a pooled average of all the class variances and for that matter the given SNR does not hold for any one of the populations. Another alternative considered in [64] is to measure noise solely on the basis of its variance.

The definition adopted here is to base the variance of the signal on the entire picture frame and in effect lump the individual class variances that may be present in

the particular data set. The reasoning behind this approach is that long before any knowledge is available about the population structure of a data set, random noise is already added to the signal; therefore, any class-dependent definitions of SNR would be unrealistic. These considerations are primarily applicable to actual data sets. In a highly controlled simulation environment, however, some or all of the above restrictions can be relaxed. For example, noise can be added to each class in different quantities in order to observe its effects on the classifiability of one particular population.

The next question to be resolved is the location, in the MSS spatial model, at which this definition of SNR applies. In Fig. 3-1 additive noise could enter both at the input and the output of the scanner system. While this is a realistic model, from a practical point of view the input noise does not limit the system performance so much due to the following reasons. First is the fact that other noise sources involved; i.e., quantization and detector noise are generally more dominant than any other disturbance arising from the atmosphere during normal operating conditions. Second, and more importantly, is the MSS response to a white noise random process. It has been pointed out that the variance of the output process is proportional to the input adjacent pixel correlation. The variation of W_s vs. ρ indicates that when the input scene displays little spatial

correlation, the variance of the output process is a very small fraction of the corresponding input quantity. Let $f(x,y)$, $N_f(x,y)$, $f'(x,y)$ and $N_f'(x,y)$ be the input random process, input additive white noise, the output random process and the noise component of the output signal respectively, then

$$f'(x,y) = f(x,y) * h(x,y) \quad (3-29)$$

$$N_f'(x,y) = N_f(x,y) * h(x,y) \quad (3-30)$$

Define $(\text{SNR})_f = \text{Var}\{f(x,y)\} / \text{Var}\{N_f(x,y)\}$ and $(\text{SNR})_f' = \text{Var}\{f'(x,y)\} / \text{Var}\{N_f'(x,y)\}$. The following inequalities hold

$$\text{Var}\{f'(x,y)\} < \text{Var}\{f(x,y)\} \quad (3-31)$$

$$\text{Var}\{N_f'(x,y)\} \ll \text{Var}\{N_f(x,y)\} \quad (3-32)$$

hence

$$(\text{SNR})_f' \gg (\text{SNR})_f \quad (3-33)$$

It then follows that the noise component of the output process (prior to quantization and detector noise) is quite negligible and for all practical purposes can be neglected. Random noise generated in the detection stage of the incoming signal is, therefore, the major disturbance factor. Having narrowed the noise contribution to one source, the logical definition of SNR would be the ratio of MSS output variance (negligible noise content) to that of quantization and detector noise (N_{dq}); i.e.,

$$(\text{SNR})_o = \text{Var}\{f'(x,y)\} / \text{Var}\{N_{dq}(x,y)\} \quad (3-34)$$

Note that for a fixed noise power, $(\text{SNR})_o$ is always smaller than $(\text{SNR})_i$, the input signal-to-noise ratio

$$(\text{SNR})_i = \text{Var}\{f(x,y)\} / \text{Var}\{N_{dq}(x,y)\} \quad (3-35)$$

Since

$$\text{Var}\{f'(x,y)\} = W_s \text{Var}\{f(x,y)\} < \text{Var}\{f(x,y)\} \quad (3-36)$$

hence

$$(\text{SNR})_o = W_s (\text{SNR})_i \quad (3-37)$$

ORIGINAL PAGE IS
OF POOR QUALITY

CHAPTER 4

Experimental Evaluation of the Parametric Multiclass Bayes Error Estimator

An experimental investigation was carried out to confirm the proper operation of the CSP error estimation algorithm described in Chapter 2. In order to satisfactorily accomplish the task, as much peripheral uncertainty as possible must be eliminated so that any deviation from the desired result can be traced directly to the methodology or the computer codes. This requirement eliminates the use of real data which is likely to have characteristics that are highly dependent on outside and generally uncontrollable elements. A more satisfactory approach is the generation of a completely synthetic data base with known and prescribed properties. After the validation process has been successfully completed, actual Landsat data will be employed and the probability of correct classification for the various populations within that set estimated by a count estimator and the CSP estimation technique and the results compared.

4.1 Description of the Data Base

The generation of a synthetic data base requires control of two characteristics; spectral and spatial. Stage I simulates M populations with N features each of which has a specified multivariate normal density function.

Let $\underline{\mu}$ and $\underline{\Sigma}$ be the desired mean and covariance matrix.

The following linear transformation on a random vector

$\underline{X} \sim N(\underline{0}, \underline{I})$ produces $\underline{Y} \sim N(\underline{\mu}, \underline{\Sigma})$

$$\underline{Y} = \underline{A} \underline{X} + \underline{\mu}$$

where $\underline{A}$ is the square-root matrix associated with $\underline{\Sigma}$, i.e.,

$$\underline{A}^T \underline{A} = \underline{\Sigma}$$

the number of samples per class is generally decided by the examination of histograms as a check for normality of the statistics. No attempt was made to incorporate the geometrical shape as a factor in generating the random field and any specified number of lines and columns in a rectangular array of points can be produced. Statistically, this data set represents an 'ideal' data set except for the lack of any serial correlation in $\underline{Y}$ caused by the same property in $\underline{X}$. The almost zero pixel-to-pixel correlation is immaterial due to the fact that the Bayes spectral classifier and the CSP error estimators do not utilize any spatial correlation information available for the data. A schematic diagram of the entire data base simulation and

model evaluation is shown in Fig. 4-1.

4.2 CSP Error Estimation Model: Validation and Checkout

With the probability of correct classification of the various populations in a data set as the prime performance index, the M-class, N-feature Bayes error estimator developed in Chapter 2 comprises the basic tool by which the MSS system model is analyzed. A comprehensive set of test procedures is required to verify the proper operation of this algorithm and to observe its response to variable operating states.

The merits of a simulated data base were discussed in sec. 4.1. The question raised now is how to select the features associated with such a base. In addressing this question, the following should be kept in mind. The main purpose here is the validation of the error estimation model independently of other system components. Therefore, the test populations need not and, indeed, cannot be 'representative' of the classes found in the multispectral data. Hence, any conclusion drawn from the results serves only to evaluate the performance of the algorithm. In generating the simulated data, however, certain general guidelines were followed.

1. The minimum number of populations should be 3 and the minimum number of dimensions preferably be the same.

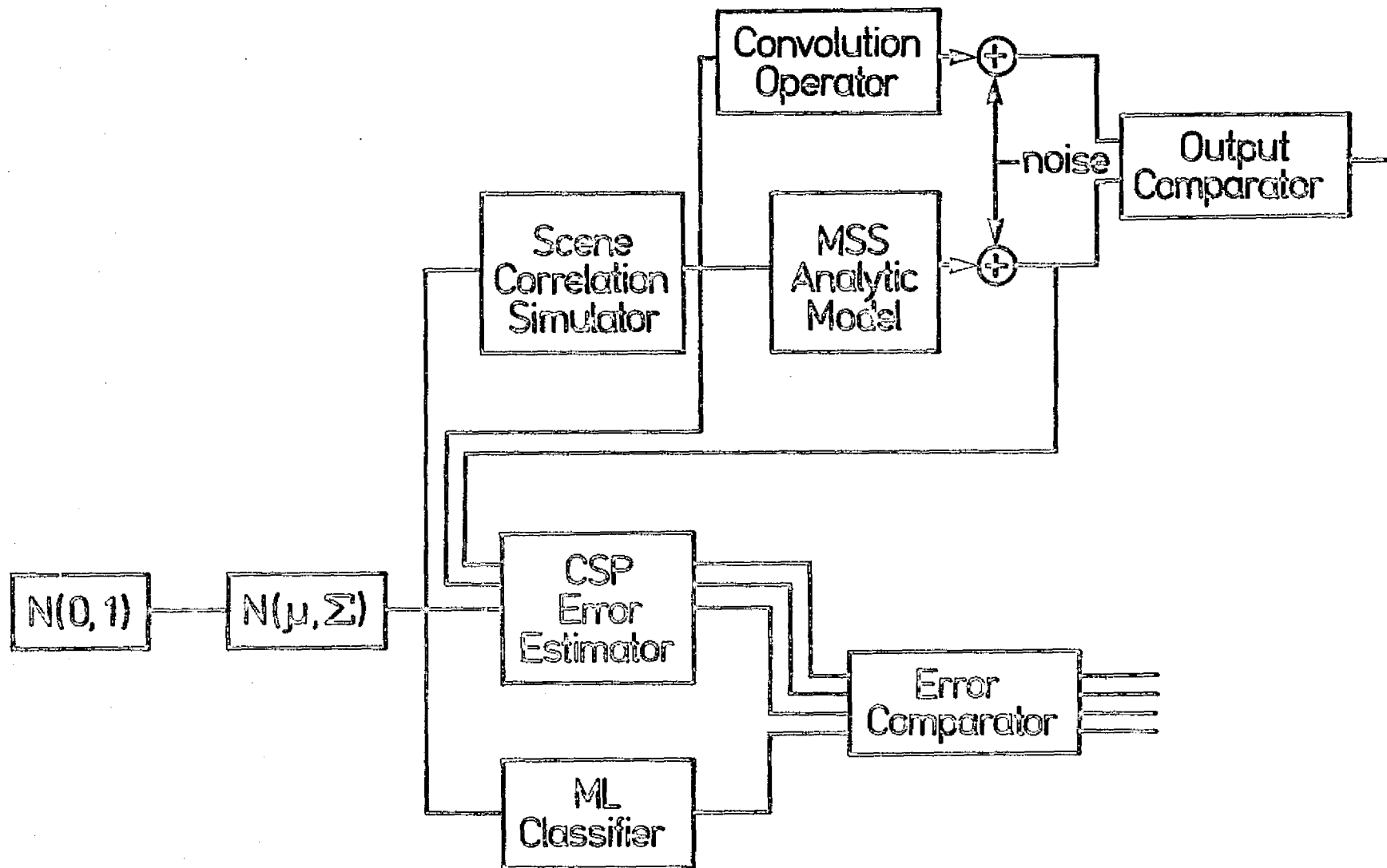


Fig. 4-1 The Block Diagram of the entire MSS Simulation Procedure.

2. The structure of the class statistics should lend itself to logical and simple manipulation of its parameters.
3. A separability measure should be defined that would reflect the changes in population parameters.

As an initial condition, three classes arranged in a simplex are considered, Fig. 4-2. This arrangement keeps the computation time low thus allowing the examination of the algorithm's performance for fine sampling grids, allows systematic parameter variation by assigning the mean vectors to different coordinates along their respective feature axis and maintains a geometrical insight as the population statistical structure is varied. Two basic categories are considered: (a) constant covariance matrices, variable mean vectors; and (b) constant mean vectors, variable covariance matrices. Because of the multiplicity of parameters describing the class statistics, it would be desirable to have a separability criterion that would lump all of the variables together and generate a single number after each change.

There are a number of separability measures to choose from. Bhattacharyya distance (B-distance) and divergence are the most notable. The former criterion will be adopted here mainly because it provides an upper bound for the error probability which can be compared with other error estimators examined here. Let the two populations ω_1 and ω_2 be distributed according to $N(\underline{\mu}_1, \underline{\Sigma}_1)$ and $N(\underline{\mu}_2, \underline{\Sigma}_2)$. Then J , the

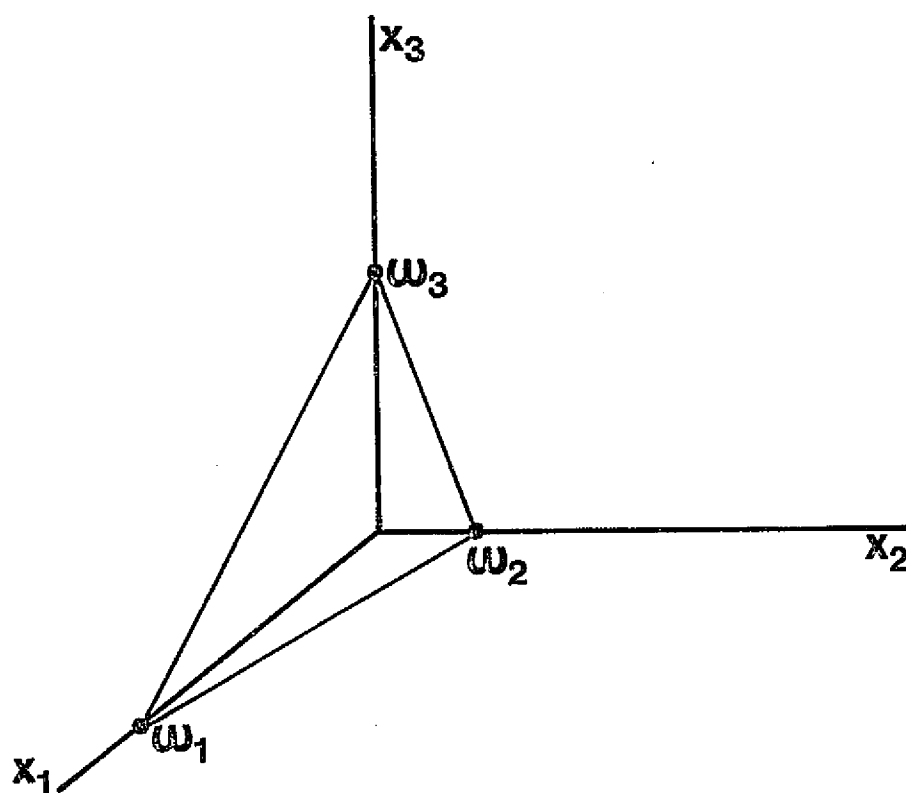


Fig. 4-2 The Configuration of Test Case Mean Vectors Arranged in a Simplex.

B-distance, and P_e , the resulting upper bound on the probability of error, are given by [41].

$$J = \frac{1}{8} (\underline{\mu}_1 - \underline{\mu}_2)^T \left(\frac{\underline{\Sigma}_1 + \underline{\Sigma}_2}{2} \right)^{-1} (\underline{\mu}_1 - \underline{\mu}_2) + \frac{1}{2} \ln \frac{|\frac{1}{2}(\underline{\Sigma}_1 + \underline{\Sigma}_2)|}{|\underline{\Sigma}_1|^{\frac{1}{2}} |\underline{\Sigma}_2|^{\frac{1}{2}}} \quad (4-3)$$

$$P_E < \sqrt{P(\omega_1)P(\omega_2)} e^{-J} = C_B \quad (4-4)$$

The Chernoff bound, C_B , in (4-4) applies to a binary set but it can be generalized by a pairwise summation. The resulting bound, however, is not adequately tight.

The only practical reference against which the results of the CSP error estimation algorithm can be compared is the Monte-Carlo (MC) type simulation of the population statistics using pseudorandom numbers, assignment of samples to their respective categories by a Bayes classifier and finally a count estimator to provide the classification accuracy. A criterion however, needs to be defined if the results of the comparison are to be meaningful. One such measure is the equality of the total number of samples used in the estimation process; i.e.,

$$\text{number of samples} | \text{MC} = (n+1)^N \quad (4-5)$$

where the right side of (4-5) is the total number of cells in the sampled space of Ω .

4.2.1 Fixed Mean, Variable Scatter

In a feature space of multivariate nature the multiplicity of dimensions generates a vast number of possible combinations of parameters to manipulate. Even for the moderate size case proposed here there are 3 variances, 9 covariances and 3 mean values capable of taking on a continuum of an infinite number of states. Therefore, a certain degree of arbitrariness must be employed in selecting the initial values and their subsequent variations. The approach selected here is the adoption of one variable statistic against a fixed background in the form of two static populations. The fixed statistic is selected after examination of the correlation matrices obtained for different types of ground cover, [65]. An attempt was made to choose correlation structures that would approximately represent two typical cases, albeit crudely. As pointed out before, whether this is true or not has little bearing on the results of this validation procedure. This choice simply displays an attempt to be as realistic as possible. Assuming that the set of three spectral bands is composed of two in the visible and one in the near-infrared, the fixed correlation matrices are given by,

$$\underline{s}_{f_2} = \begin{bmatrix} 1 & 0.8 & 0 \\ & 1 & 0.1 \\ & & 1 \end{bmatrix}$$

(4-6)

$$\underline{s}_{f_3} = \begin{bmatrix} 1 & 0.94 & 0.15 \\ & 1 & 0.05 \\ & & 1 \end{bmatrix}$$

The class with the variable scatter is specified by a choice of 4 different across-band correlation values ranging from a low of 0.15, medium of 0.45, medium high of 0.75 and a high of 0.95. The permutation of these four numbers taken 3 at a time generates 24 different cases out of which 13 result in invalid non-positive definite matrices. For each remaining case, an average B-distance $\bar{J}$ is computed and the 11 permissible combinations are tabulated in the order of increasing separability, Table 4-1. $\bar{J}\%$ is the value of $\bar{J}$ normalized to the highest $\bar{J}$ in the table and s_{ij} is the channel i and j correlation coefficient. The means are fixed at 0.7σ on each axis with $\sigma=1$. The grid size for the CSP error estimation technique ranged from 4 to 14 cells per axis with an increment of 1 which is equivalent to 4^3 to 14^3 samples for the corresponding MC estimator. For each of the 11 cases outlined in Table 4-1, there exists 3 plots. The first two show the variation of the CSP (MC) error estimator vs. grid (sample) size and the

ORIGINAL PAGE IS
OF POOR QUALITY

third plot shows the variance of the error estimate for the two aforementioned techniques. Each plot is accompanied by a table of values. Throughout sec. 4.2.1, 'case i' corresponds to the particular separability of rank i (from the top) of Table 4-1 and $N_B\%$ is the number of boundary cells as a percentage of the inside cells, and G_s , the grid size is the number of cells per axis.

The results of the variable scatter geometry provide the basic understanding of the potentials and operating principles of this error estimation technique and exhibit many properties universal to this algorithm. The first and probably the most important item to be explored is the variation and dependence of the estimate on the grid size. This relationship is particularly crucial due to the fact that although there is a theoretical convergence established, the rate of convergence determines the feasibility of implementation of this technique as a viable alternative to other data dependent algorithms. This is especially true since the number of cells within the grid bears an exponential relationship with the dimensionality of the data. Examination of the CSP error estimator vs. grid size plots quickly disposes of this concern. The pattern exhibited throughout is one of a rapid climb to a steady state value and oscillations of small magnitude around it. The rapid convergence is best demonstrated in Case 6. Where the estimate of the overall classification accuracy

at the smallest grid size was off 7.1% from its final value, it jumped 6.2% by incrementing the grid size by one step to 5 cells per axis and from then on gained only 0.9% to level off at 72.9% for 14 cells per axis. In terms of the total number of cells involved, the initial rise of 6.2% was gained by an increase of 61 cells while the addition of 2619 more cells improved the estimate by only 0.9%. Similar behavior is observed in Case 2 where the one step rise of 7.8% was accompanied by a 10 step rise of 1.1%. These effects are evident in all 11 cases with varying degrees of intensity. On the average the initial rise of 5.14% was followed by a 1.78% increase toward the final value.

This property is remarkable in view of the performance of various sampling techniques. For a 3 dimensional grid with 5 cells per axis, there are a total of 125 points involved which provide an estimate of aforementioned quality. The performance of the MC technique with that small a sample size is totally inadequate. In fact, generating the required Gaussian data base with 125 samples is itself very difficult. Fig. 4-3 demonstrates the deviation from normality of the statistics for small sample size while for comparison purposes, a corresponding histogram using 2744 (14^3) samples is shown in Fig. 4-4. It is, therefore, clear that small sample behavior of the CSP technique is very superior to small sample size behavior of the Monte Carlo technique. It can be argued, however,

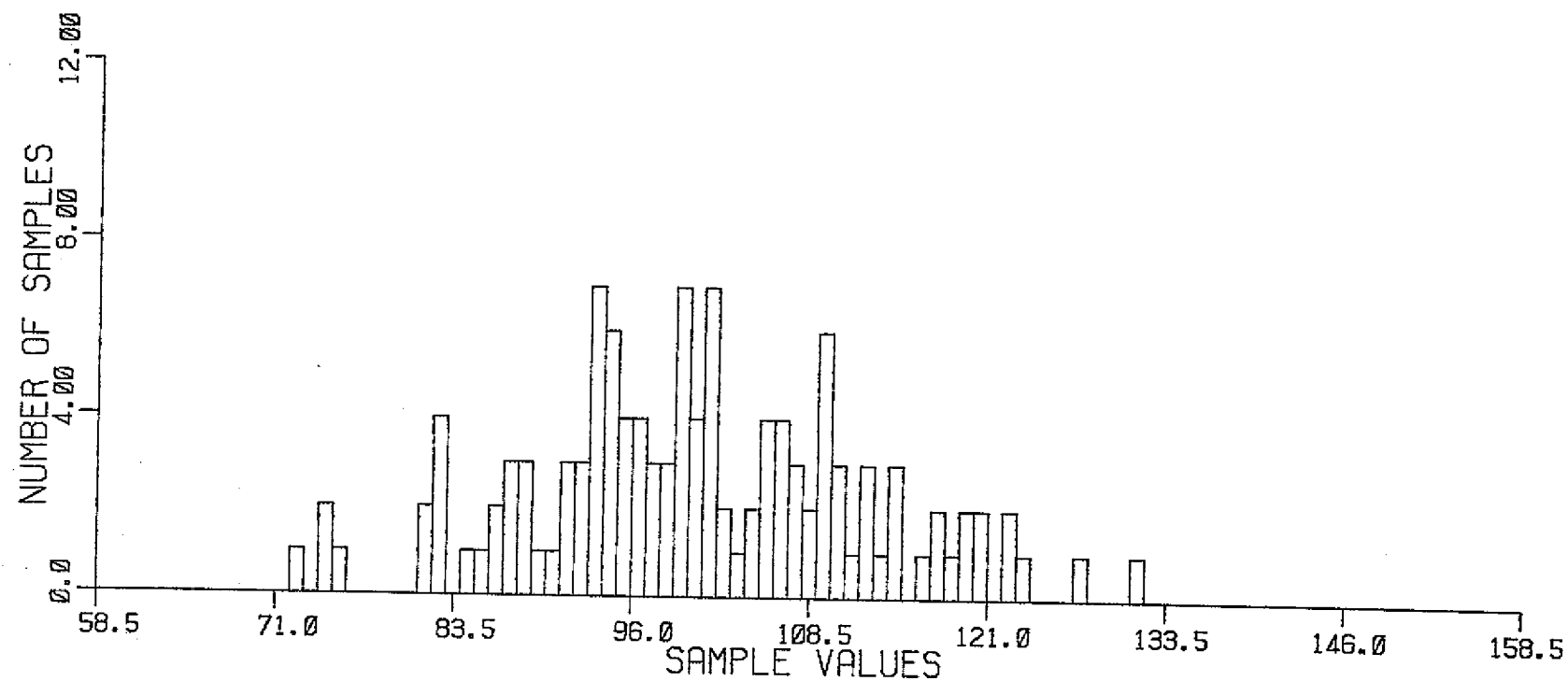


Fig. 4-3 Small Sample Size Histogram of Simulated Normal Data.

ORIGINAL PAGE IS
OF POOR QUALITY

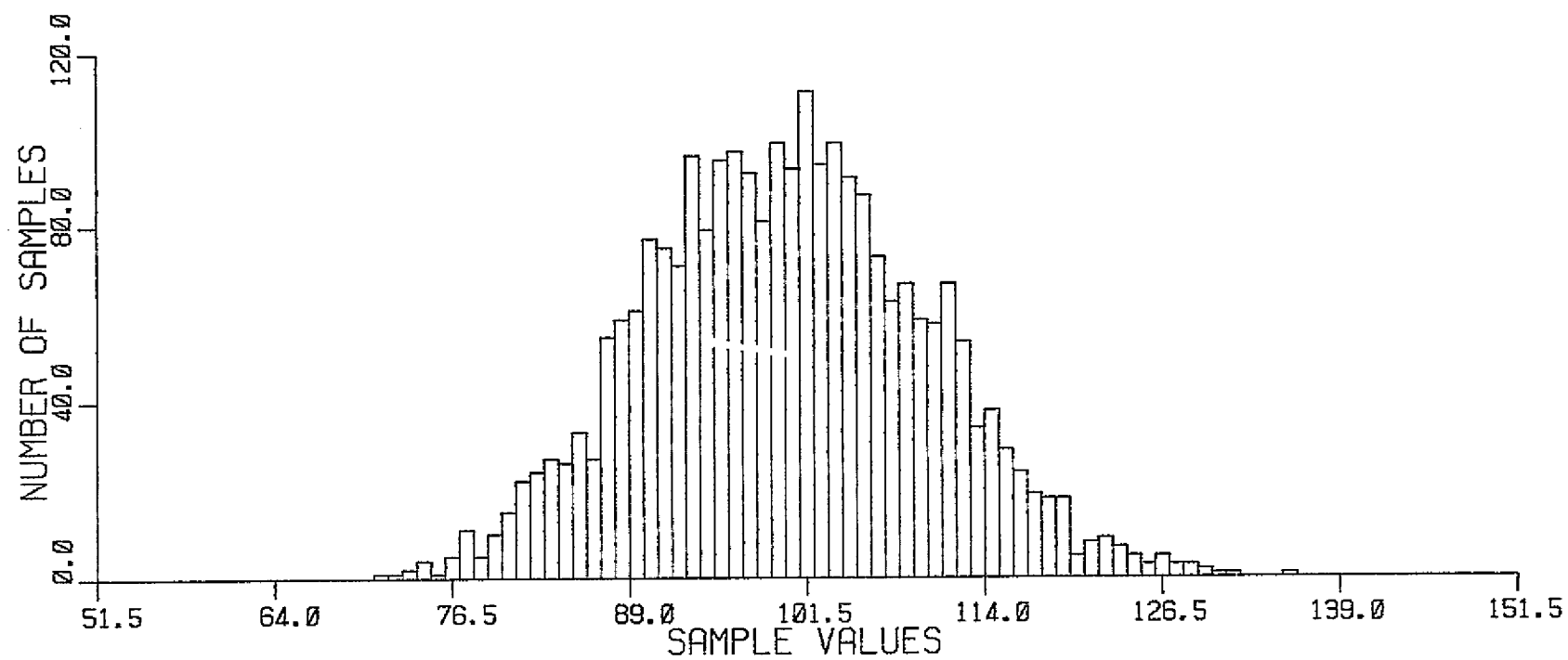


Fig. 4-4 Medium Sample Size Histogram of Simulated Normal Data

TABLE 4- 1 TEST CASES ARRANGED BY INCREASING SEPERABILITY.
VARIABLE SCATTER.

s_{12}	s_{13}	s_{23}	$\bar{J}$	$\bar{J}\%$	$C_B\%$
0.75	0.15	0.45	0.50	29	39.0
0.45	0.15	0.75	0.52	30	40.6
0.75	0.45	0.15	0.54	31	41.6
0.15	0.45	0.75	0.58	34	43.9
0.45	0.75	0.15	0.59	34	44.3
0.15	0.75	0.45	0.60	35	44.7
0.45	0.15	0.95	1.48	86	69.4
0.95	0.15	0.45	1.52	88	69.5
0.45	0.95	0.15	1.58	91	70.6
0.15	0.95	0.45	1.58	92	70.7
0.95	0.45	0.15	1.72	100	72.1

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 4- 2 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 1

G_S	$\hat{P}_{C \omega_1}$		$\hat{P}_{C \omega_2}$		$\hat{P}_{C \omega_3}$		$\hat{P}_C$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
4	54.9	71.9	69.6	75.0	70.0	81.3	64.8	76.0
5	64.8	68.6	72.2	74.4	73.3	81.0	70.1	74.7
6	63.6	69.9	71.0	75.0	73.1	78.1	69.3	74.3
7	65.6	69.4	71.4	73.8	76.3	76.9	71.1	73.4
8	66.3	64.3	69.4	73.6	76.5	76.2	70.8	71.3
9	68.2	68.4	70.3	76.7	75.9	75.9	71.5	73.7
10	69.3	66.6	71.7	74.4	77.7	75.3	72.9	72.1
11	68.4	69.1	72.8	73.8	76.8	75.6	72.7	72.9
12	68.8	67.6	73.7	76.9	76.6	76.4	73.0	73.6
13	68.9	69.9	74.2	74.2	76.6	77.4	73.2	73.8
14	68.6	69.3	74.3	74.1	75.9	76.7	72.9	73.4

TABLE 4- 3 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

G_S	CSP	MC	N_B^2
4	4.3	5.7	99.9
5	2.6	4.1	64.4
6	3.0	3.1	68.8
7	2.6	2.5	51.3
8	2.2	2.0	52.8
9	1.9	1.7	41.5
10	1.7	1.4	41.5
11	1.5	1.3	34.2
12	1.4	1.1	35.3
13	1.3	1.0	30.0
14	1.2	0.9	29.8

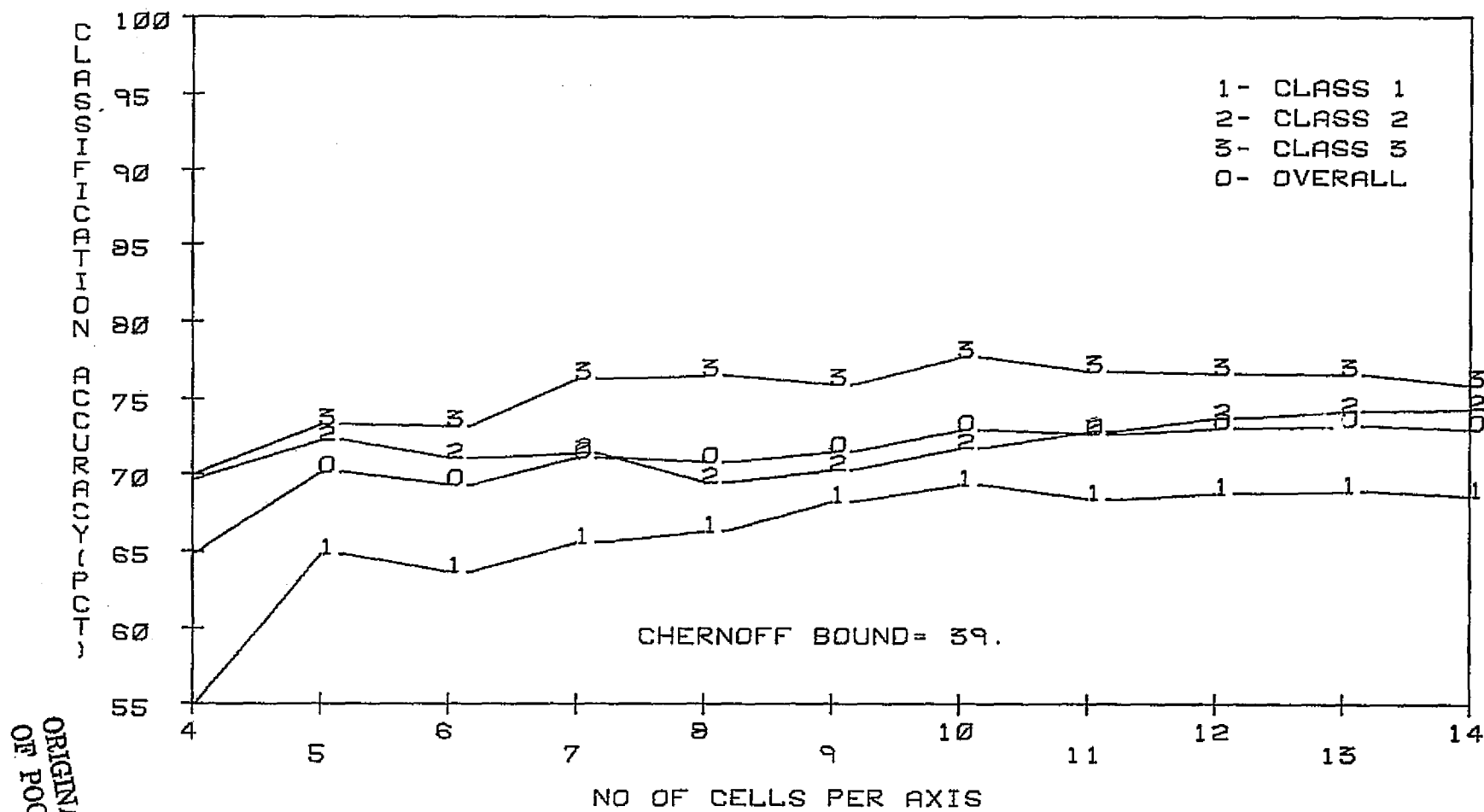


FIG. 4-5 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 1

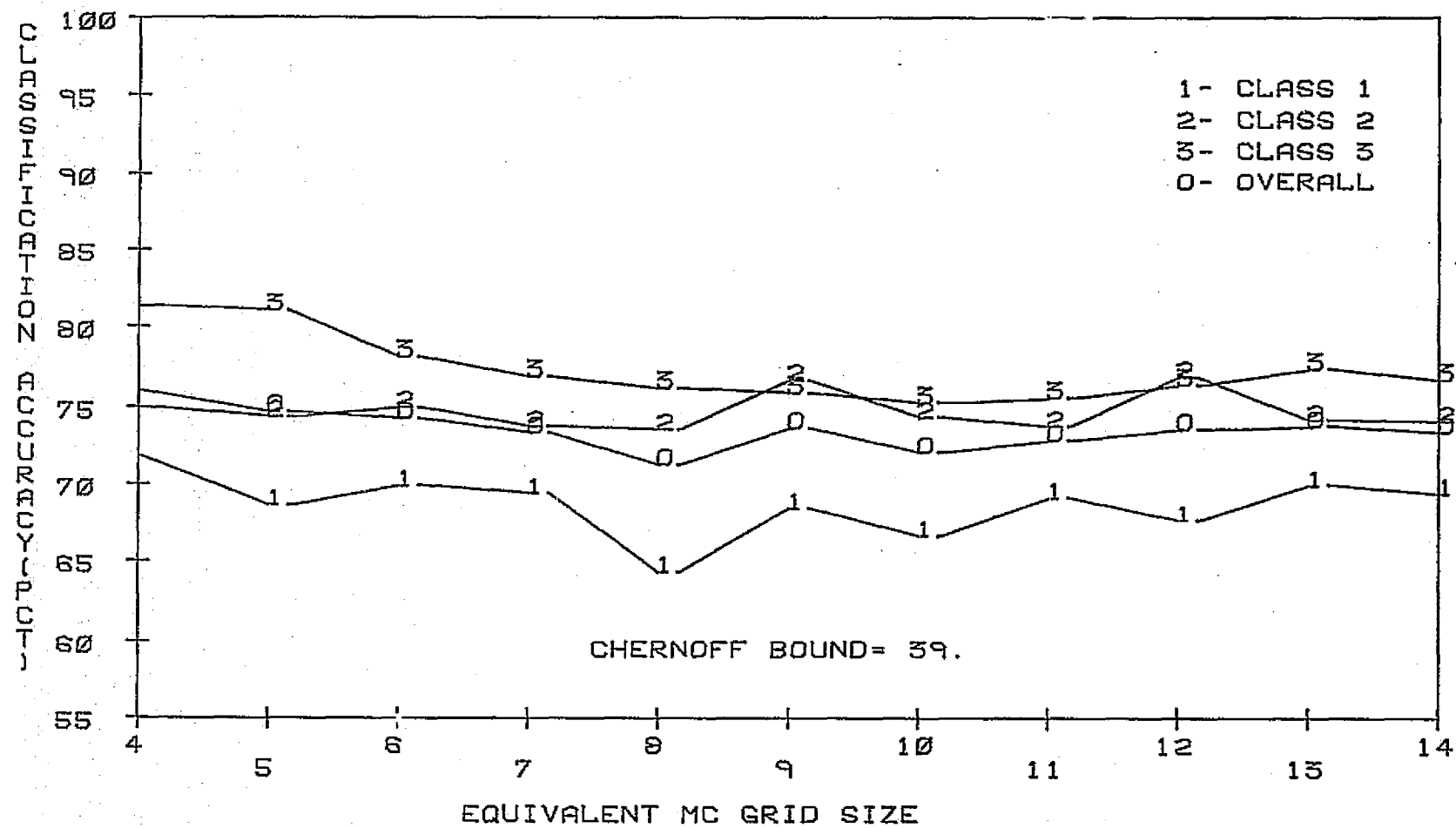


FIG. 4-6 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 1

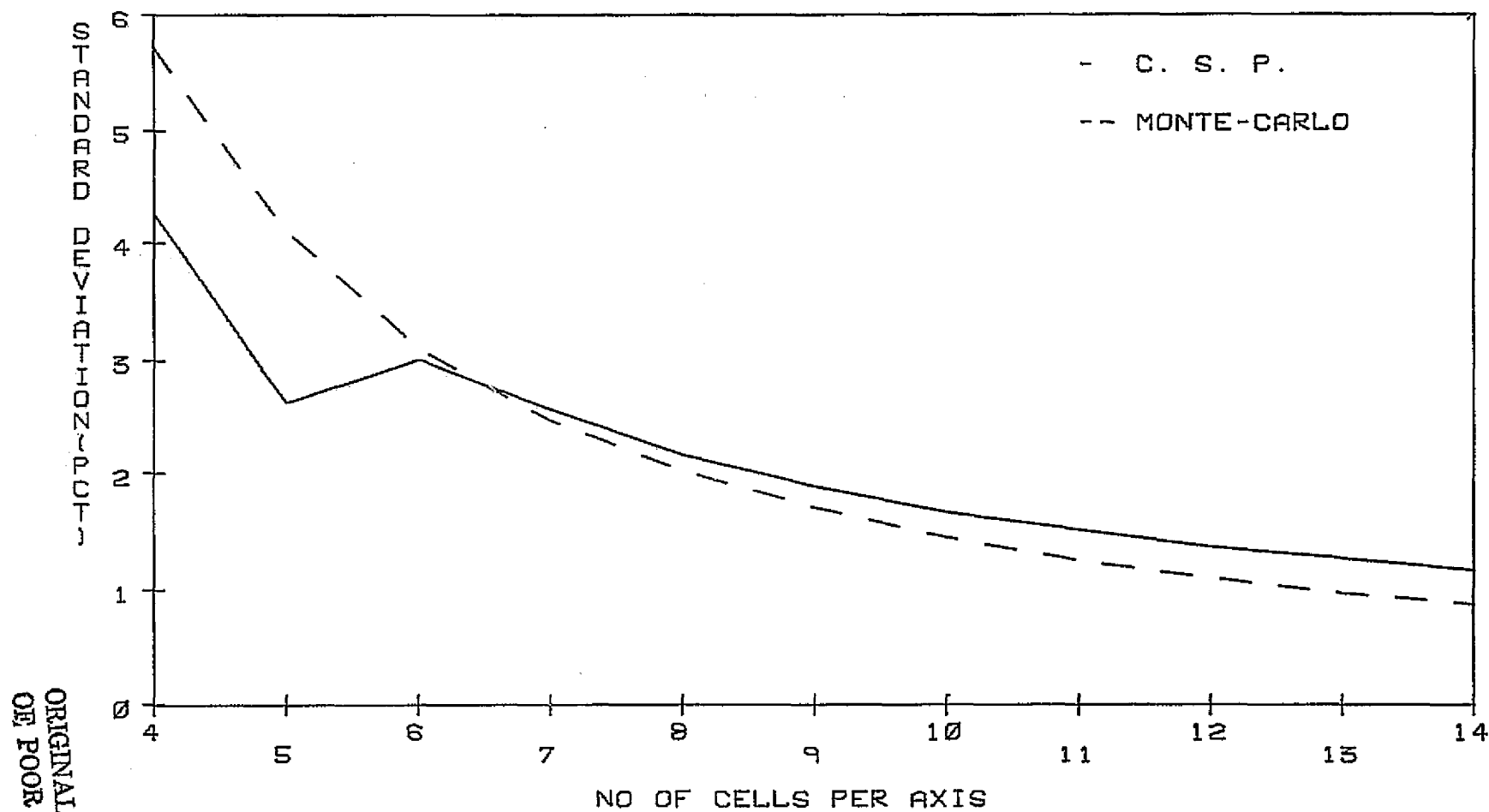


FIG. 4-7 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 1

TABLE 4- 4 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 2

G_S	$\hat{P}_{C \omega_1}$		$\hat{P}_{C \omega_2}$		$\hat{P}_{C \omega_3}$		$\hat{P}_C$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
4	48.7	68.8	68.8	75.0	71.4	89.1	63.0	77.6
5	67.2	67.8	71.7	71.9	73.6	80.2	70.8	73.3
6	61.1	65.3	70.9	73.5	75.8	81.6	69.3	73.5
7	65.3	66.7	71.4	74.1	77.7	78.1	71.5	72.9
8	64.7	63.4	70.3	72.5	78.4	78.1	71.2	71.3
9	65.1	65.0	70.6	76.0	78.1	77.9	71.3	73.0
10	64.3	61.6	72.1	74.0	78.5	77.7	71.7	71.1
11	63.2	66.4	72.0	73.8	78.4	77.8	71.2	72.7
12	64.1	64.7	73.1	75.1	78.3	78.0	71.8	72.6
13	64.4	66.0	73.3	73.5	78.0	79.3	71.9	72.9
14	64.5	63.6	73.4	74.1	77.8	78.5	71.9	72.1

TABLE 4- 5 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

G_S	CSP	MC	$N_B\%$
4	3.3	6.0	74.0
5	2.5	4.3	40.5
6	2.2	3.3	41.8
7	1.5	2.6	30.0
8	1.2	2.1	31.7
9	1.2	1.8	25.1
10	1.4	1.5	26.0
11	1.4	1.3	21.2
12	1.1	1.1	21.2
13	0.7	1.0	18.2
14	0.8	0.9	18.2

ORIGINAL PAGE IS
OF POOR QUALITY

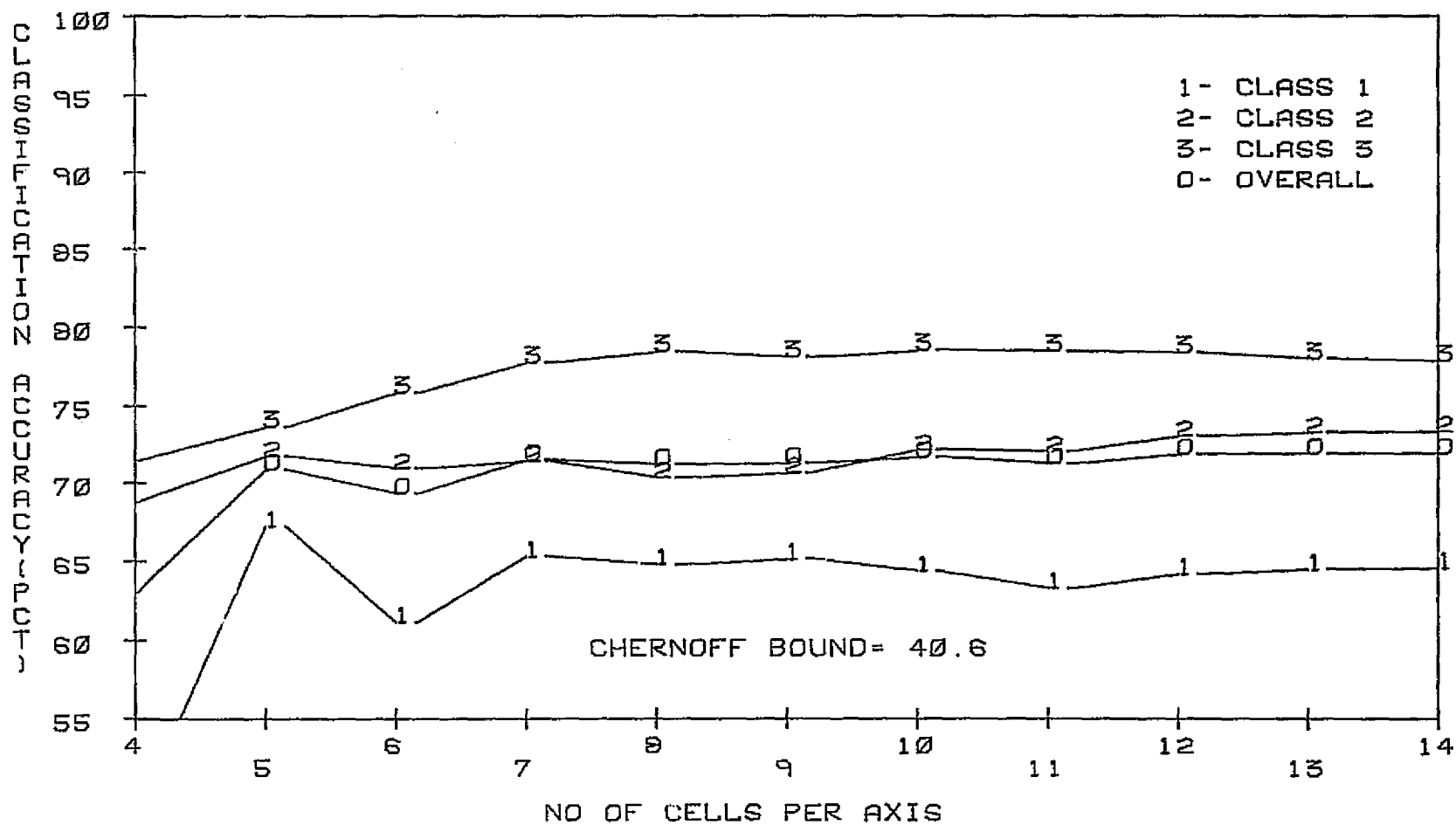


FIG. 4-8 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 2

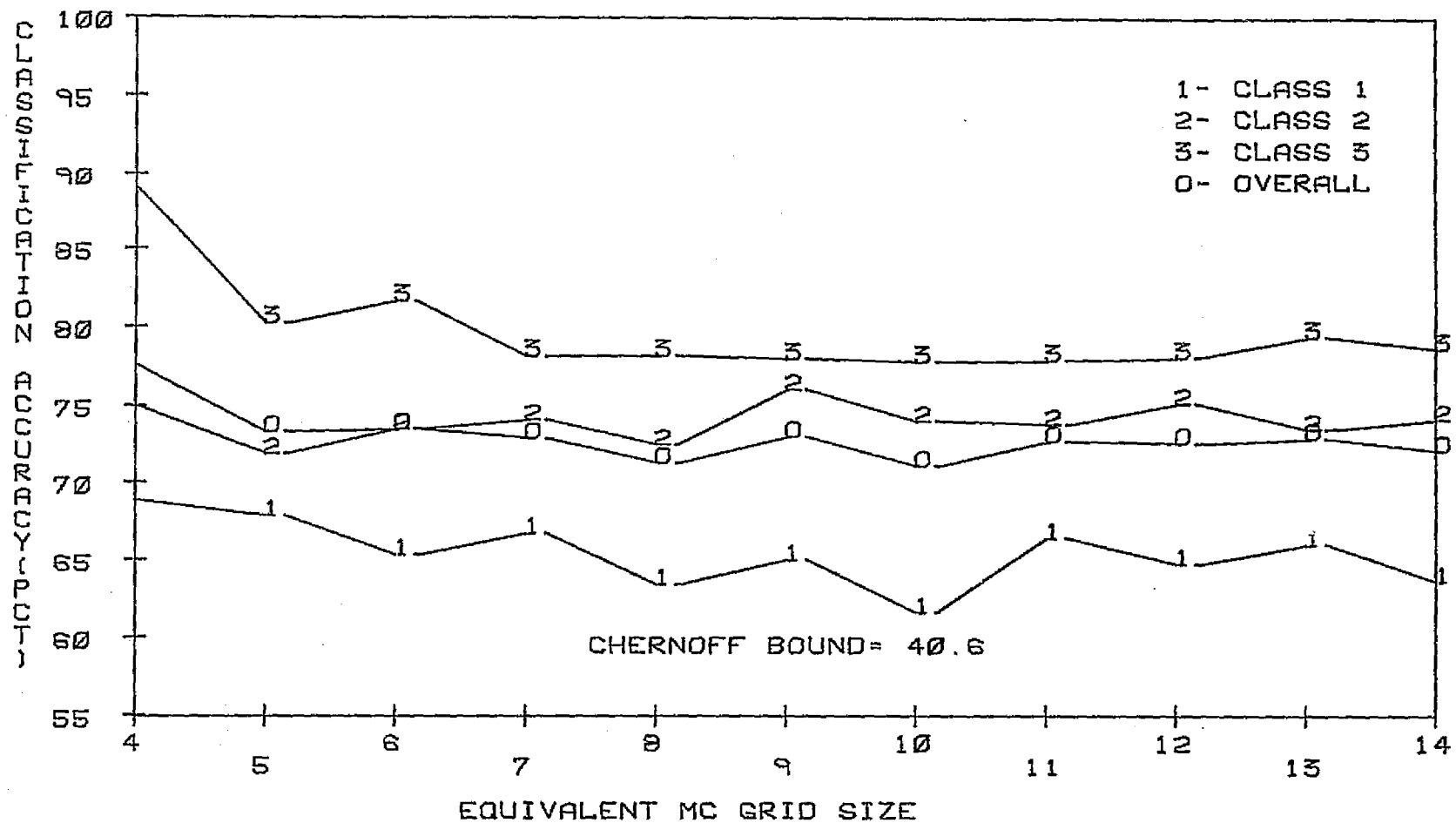


FIG. 4-9 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 2

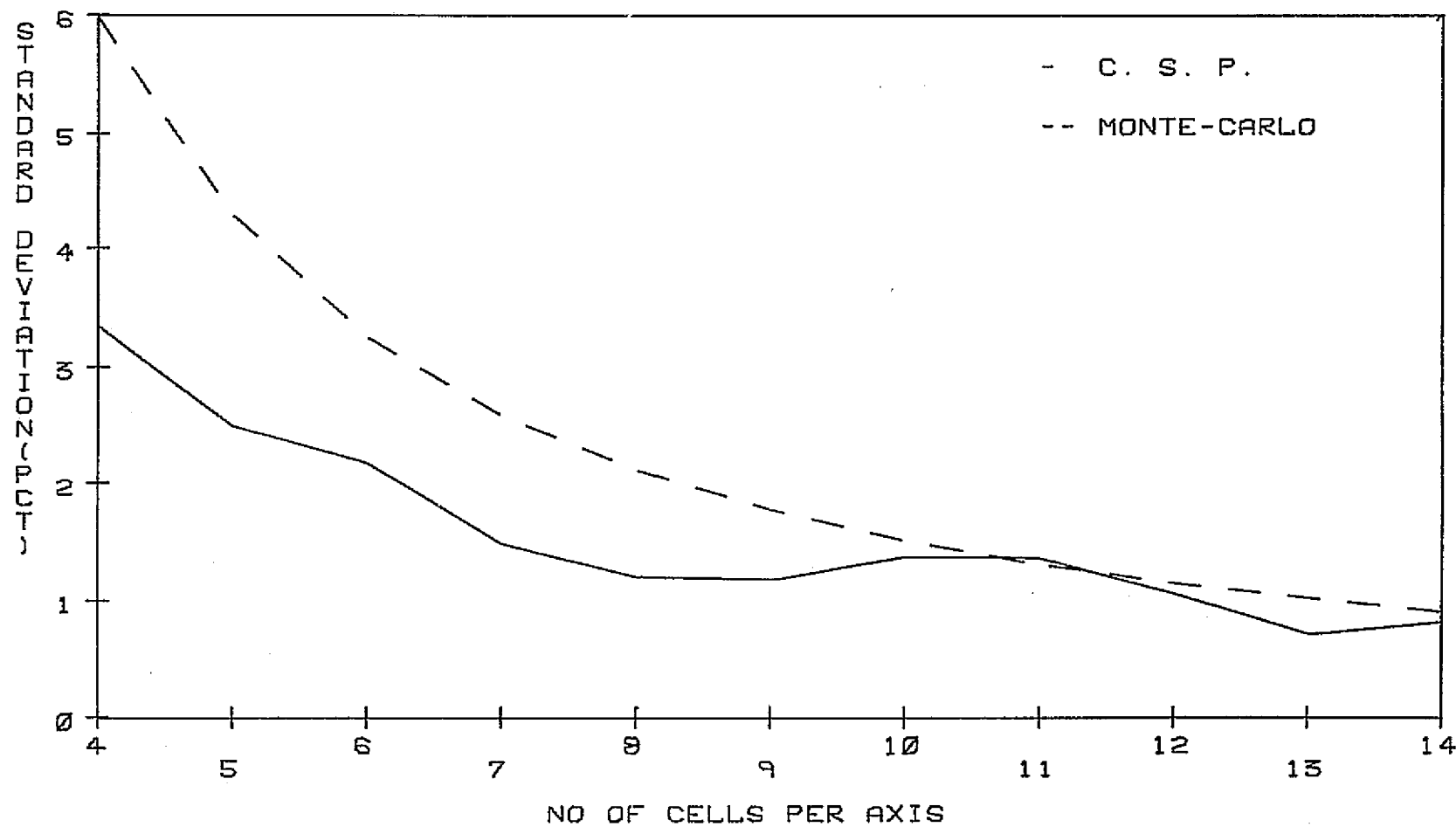


FIG. 4-10 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 2

TABLE 4- 6 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 3

G_S	$\hat{P}_{C \omega_1}$		$\hat{P}_{C \omega_2}$		$\hat{P}_{C \omega_3}$		$\hat{P}_C$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
4	62.4	73.4	65.7	73.4	72.6	82.8	66.9	76.6
5	66.8	77.7	67.5	71.9	73.7	77.7	69.4	75.8
6	68.1	78.1	67.6	70.4	73.9	78.1	69.9	75.5
7	71.5	80.9	67.8	71.6	73.8	71.2	71.0	76.5
8	73.3	73.6	67.4	70.0	75.7	77.3	72.1	73.6
9	76.0	75.9	67.8	74.5	75.5	75.6	73.1	75.3
10	79.4	74.0	69.0	70.4	76.8	75.2	75.1	73.2
11	78.6	75.2	70.2	71.5	76.3	76.2	75.0	74.3
12	78.3	75.0	71.1	73.8	76.3	76.9	75.2	75.2
13	78.8	77.1	71.7	71.9	76.7	77.1	75.8	75.4
14	78.0	75.8	71.7	70.9	76.3	75.9	75.3	74.2

TABLE 4- 7 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

G_S	CSP	MC	$N_B^{\%}$
4	4.0	5.2	91.4
5	3.4	3.7	64.1
6	3.0	2.8	58.5
7	2.6	2.2	46.4
8	2.2	1.8	43.7
9	1.4	1.5	35.4
10	1.2	1.3	34.4
11	1.1	1.1	28.7
12	1.0	1.0	27.9
13	0.9	0.9	24.0
14	0.8	0.8	23.8

ORIGINAL PAGE IS
OF POOR QUALITY

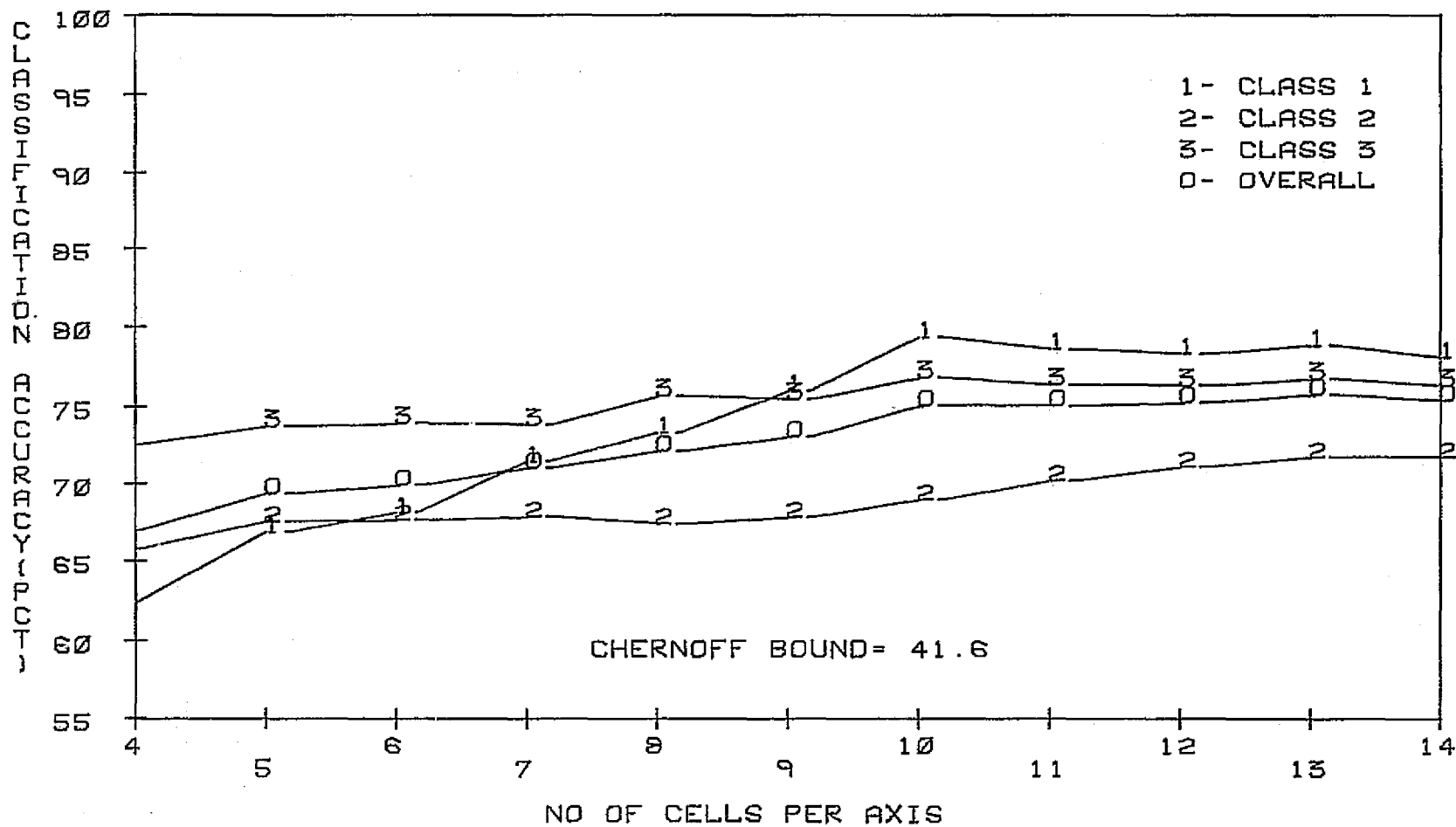


FIG. 4-11 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 3

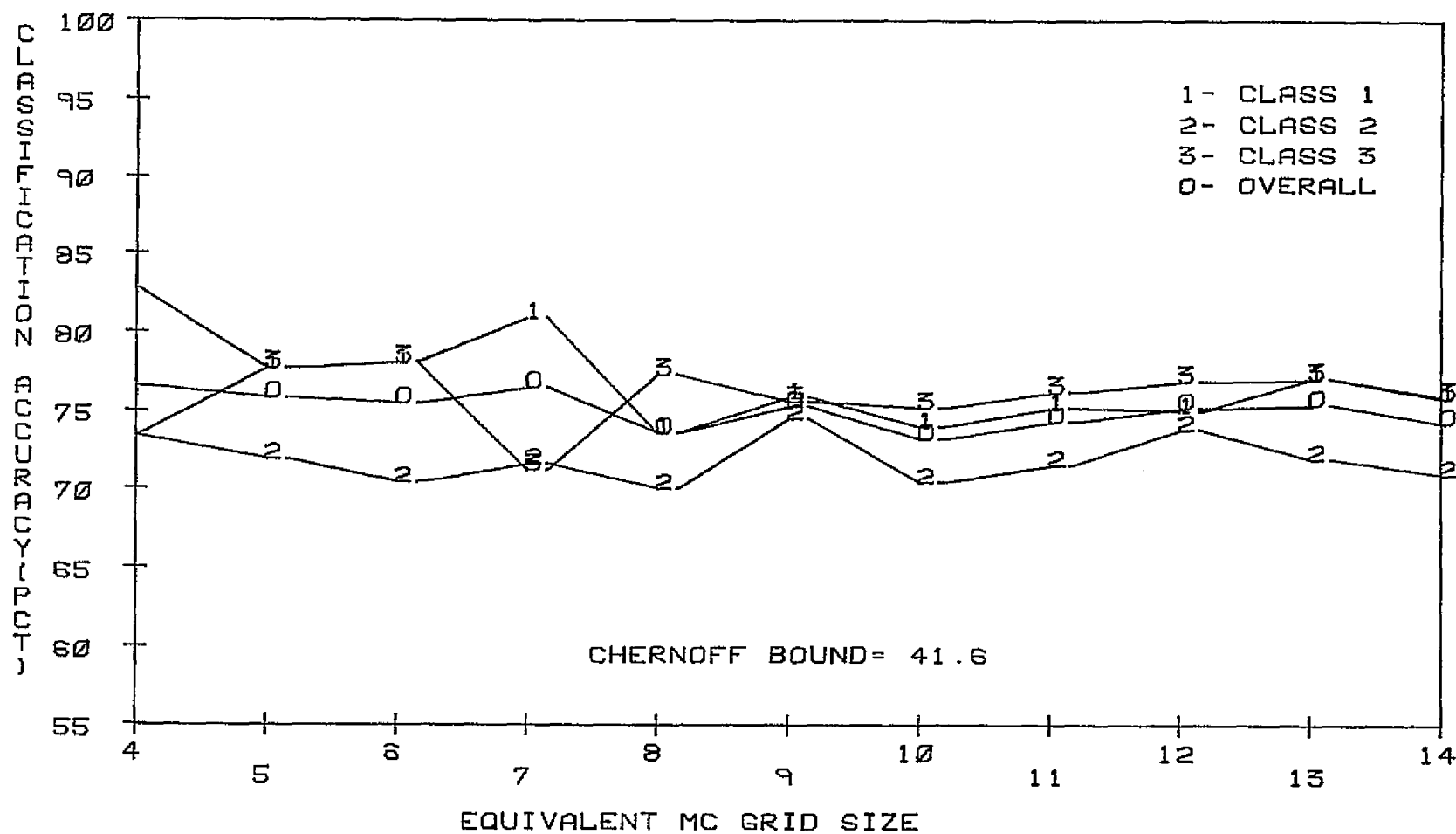


FIG. 4-12 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 3

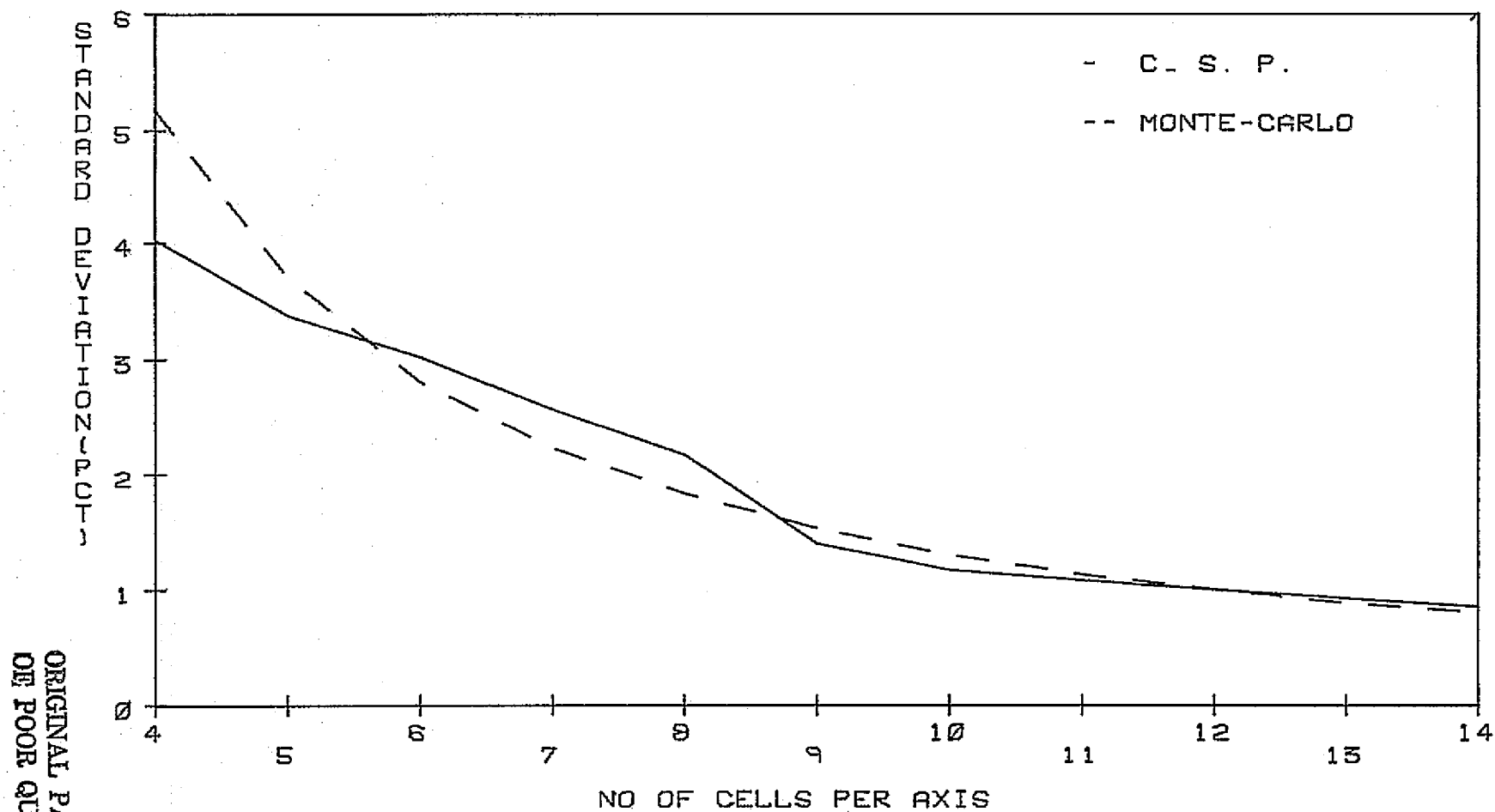


FIG. 4-13 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 3

TABLE 4- 8 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 4

G_S	$\hat{P}_{C \omega_1}$		$\hat{P}_{C \omega_2}$		$\hat{P}_{C \omega_3}$		$\hat{P}_C$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
4	58.3	67.2	68.4	71.9	73.0	90.6	66.5	76.6
5	63.8	66.9	72.3	71.9	74.5	81.0	70.2	73.3
6	62.4	66.3	70.4	73.0	75.5	81.6	69.4	73.6
7	66.1	66.0	71.1	73.5	77.9	79.3	71.7	72.9
8	65.3	65.5	69.3	71.9	78.6	78.7	71.1	72.0
9	64.8	65.3	70.4	75.3	78.0	78.7	71.1	73.1
10	64.2	62.7	71.6	72.4	78.9	77.7	71.6	71.0
11	65.6	65.6	71.4	72.4	79.2	78.3	72.1	72.1
12	67.2	65.5	72.5	74.5	79.7	78.8	73.1	72.9
13	66.6	67.1	72.6	72.8	79.7	80.5	72.9	73.5
14	66.1	64.4	73.3	73.2	79.1	79.1	72.8	72.2

TABLE 4- 9 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

G_S	CSP	MC	$N_B^{\%}$
4	2.5	5.9	48.6
5	1.7	4.2	32.0
6	1.9	3.2	27.7
7	1.2	2.5	23.4
8	1.1	2.1	23.3
9	0.7	1.8	19.0
10	0.6	1.5	18.0
11	0.5	1.3	15.6
12	0.6	1.1	15.6
13	0.5	1.0	13.9
14	0.4	0.9	13.3

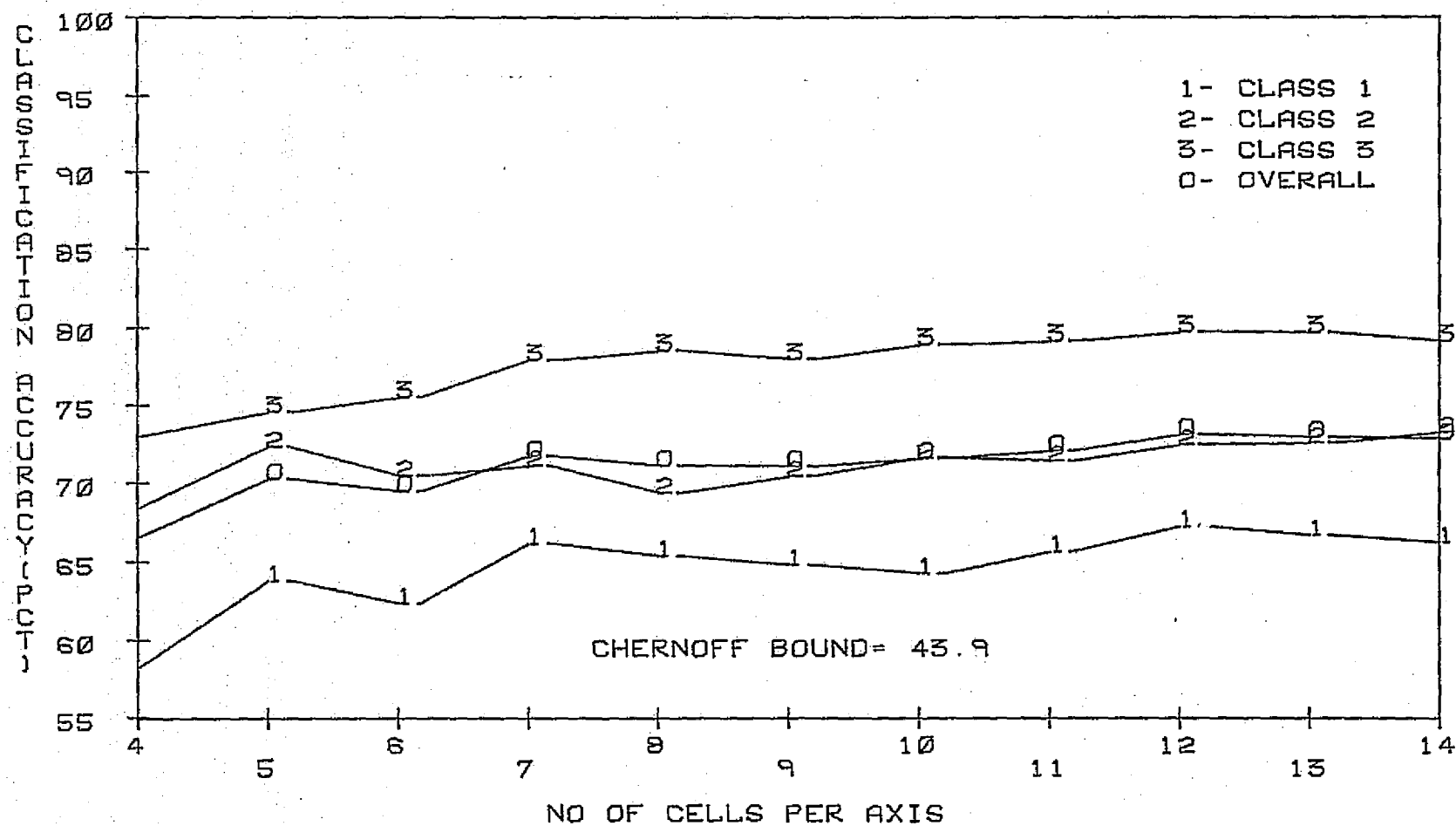


FIG. 4-14 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 4

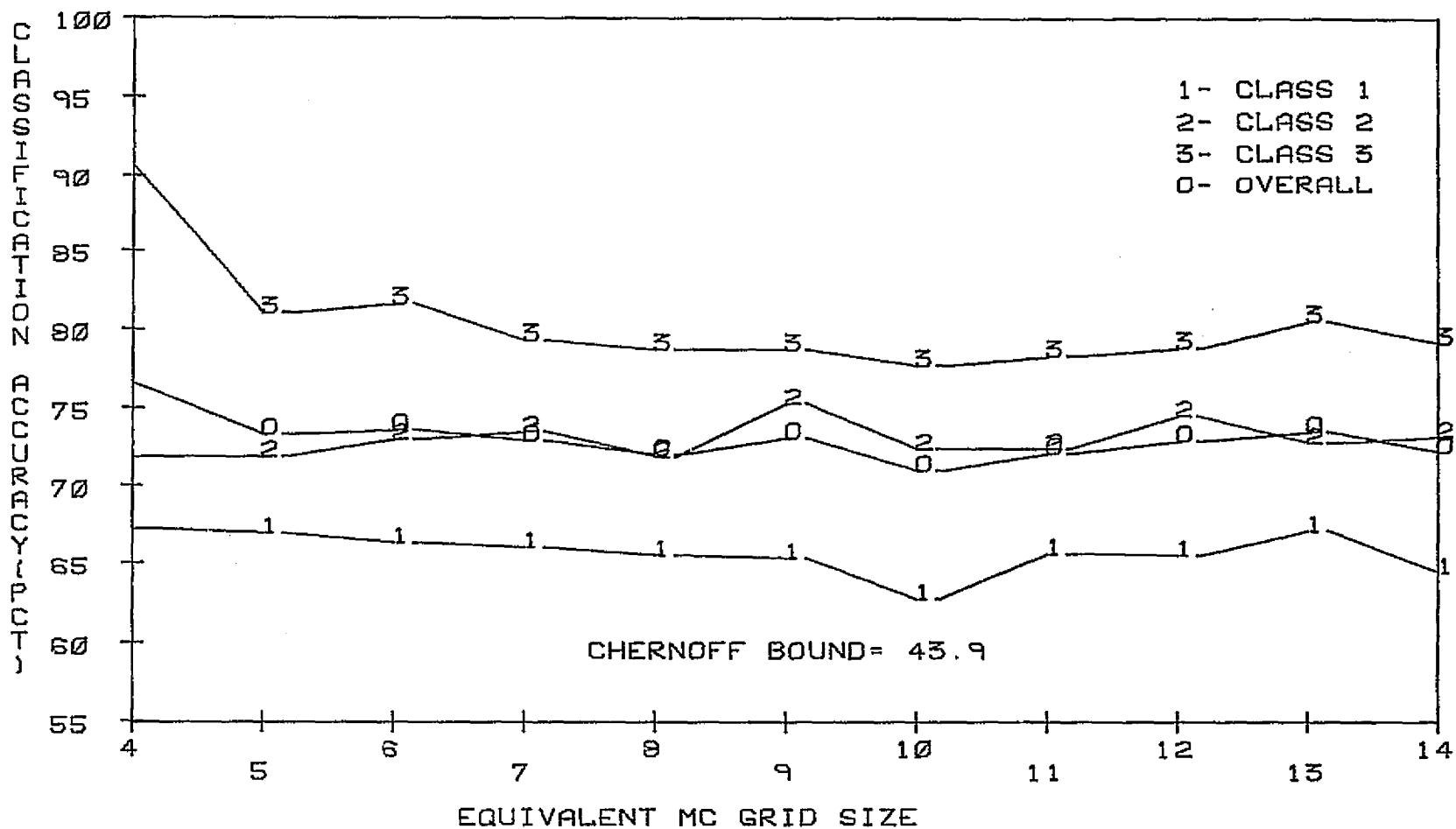


FIG. 4-15 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 4

ORIGINAL PAGE IS
OF POOR QUALITY

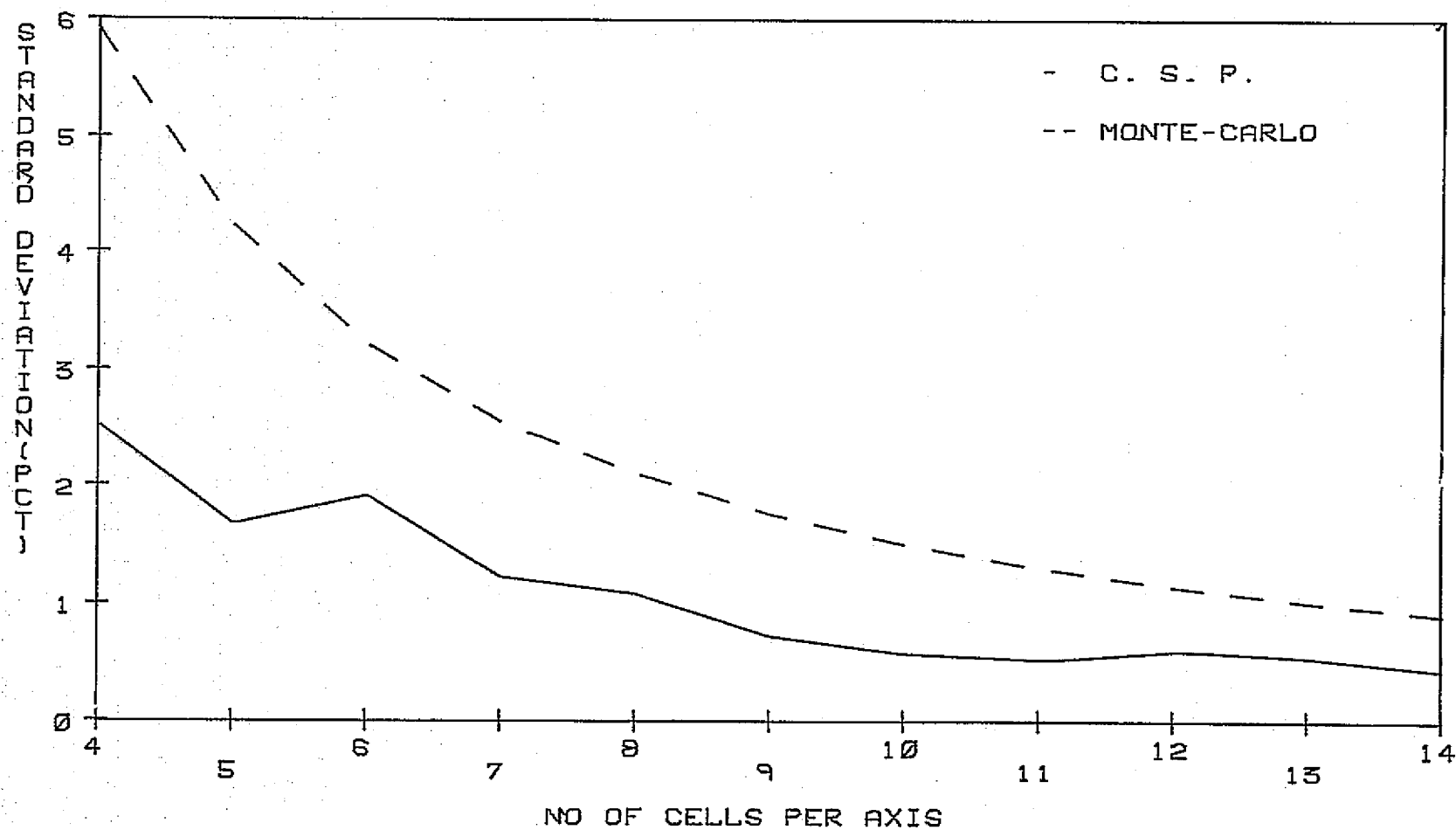


FIG. 4-16 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 4

TABLE 4-10 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 5

G_s	$\hat{P}_{c \omega_1}$		$\hat{P}_{c \omega_2}$		$\hat{P}_{c \omega_3}$		$\hat{P}_c$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
4	69.7	65.6	66.7	73.4	74.2	85.9	67.2	75.0
5	70.3	69.4	68.9	71.1	74.6	80.2	71.3	73.6
6	71.7	75.5	67.1	73.5	73.5	81.1	70.8	76.7
7	71.9	73.5	67.3	70.7	75.4	76.5	71.5	73.6
8	73.3	70.0	67.0	72.3	75.5	78.3	71.9	73.6
9	72.6	73.3	68.1	74.3	76.8	77.2	72.5	74.9
10	71.9	69.3	68.5	70.6	77.6	76.4	72.6	72.1
11	72.4	72.6	69.3	71.2	77.8	77.1	73.2	73.6
12	71.7	72.0	71.4	73.7	78.1	77.2	73.7	74.3
13	73.3	75.2	71.7	71.4	77.6	78.8	74.2	75.1
14	73.0	73.0	71.3	70.2	77.1	76.7	73.8	73.3

TABLE 4-11 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

G_s	CSP	MC	$N_B^{\%}$
4	3.7	5.5	54.5
5	3.2	4.0	43.9
6	2.6	3.0	37.0
7	2.2	2.4	30.8
8	1.6	2.0	28.3
9	1.4	1.6	24.5
10	1.3	1.4	22.6
11	1.4	1.2	19.7
12	1.3	1.1	18.3
13	1.0	0.9	17.1
14	1.0	0.8	16.4

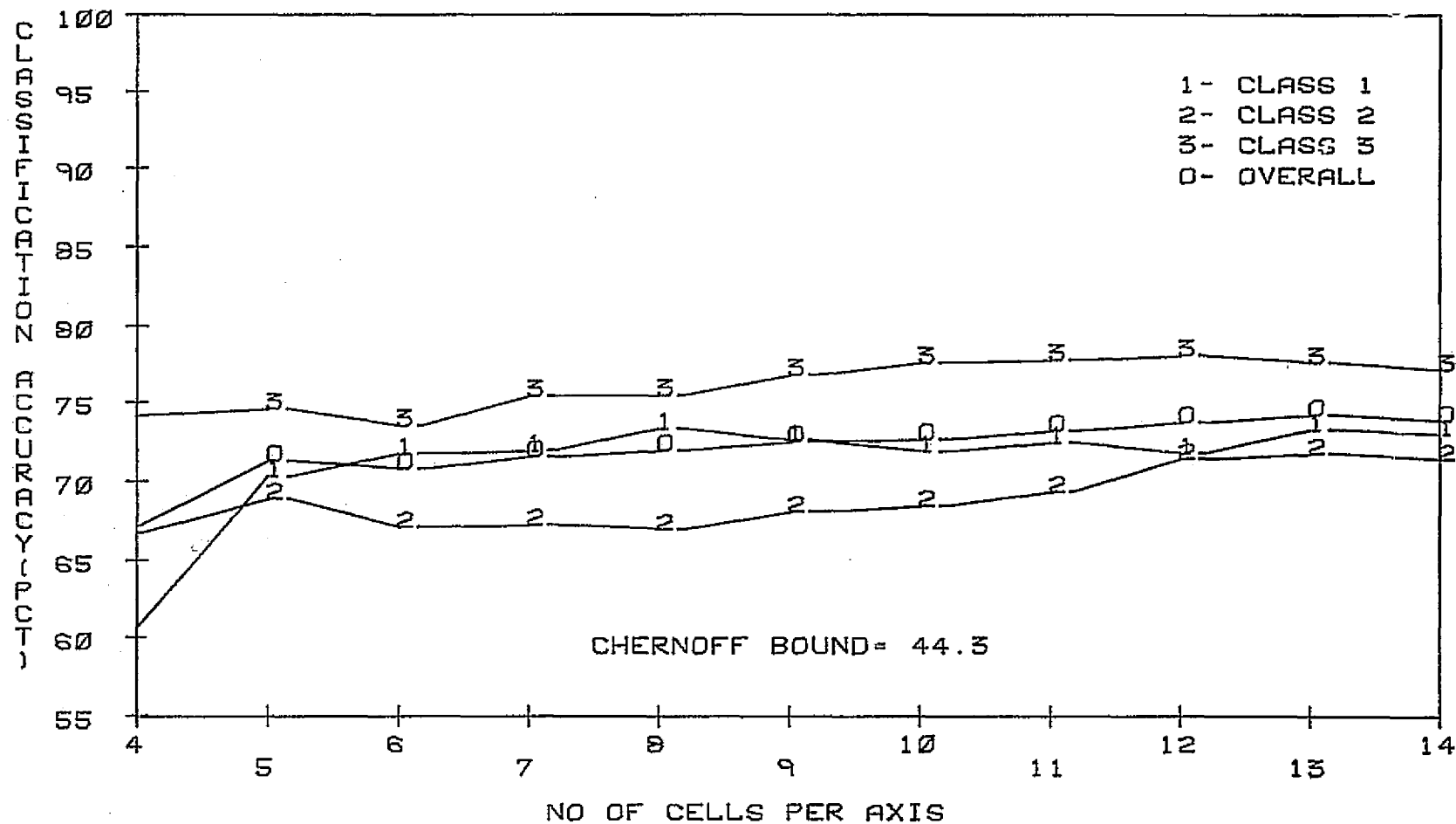


FIG. 4-17 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 5

ORIGINAL PAGE IS
OF POOR QUALITY

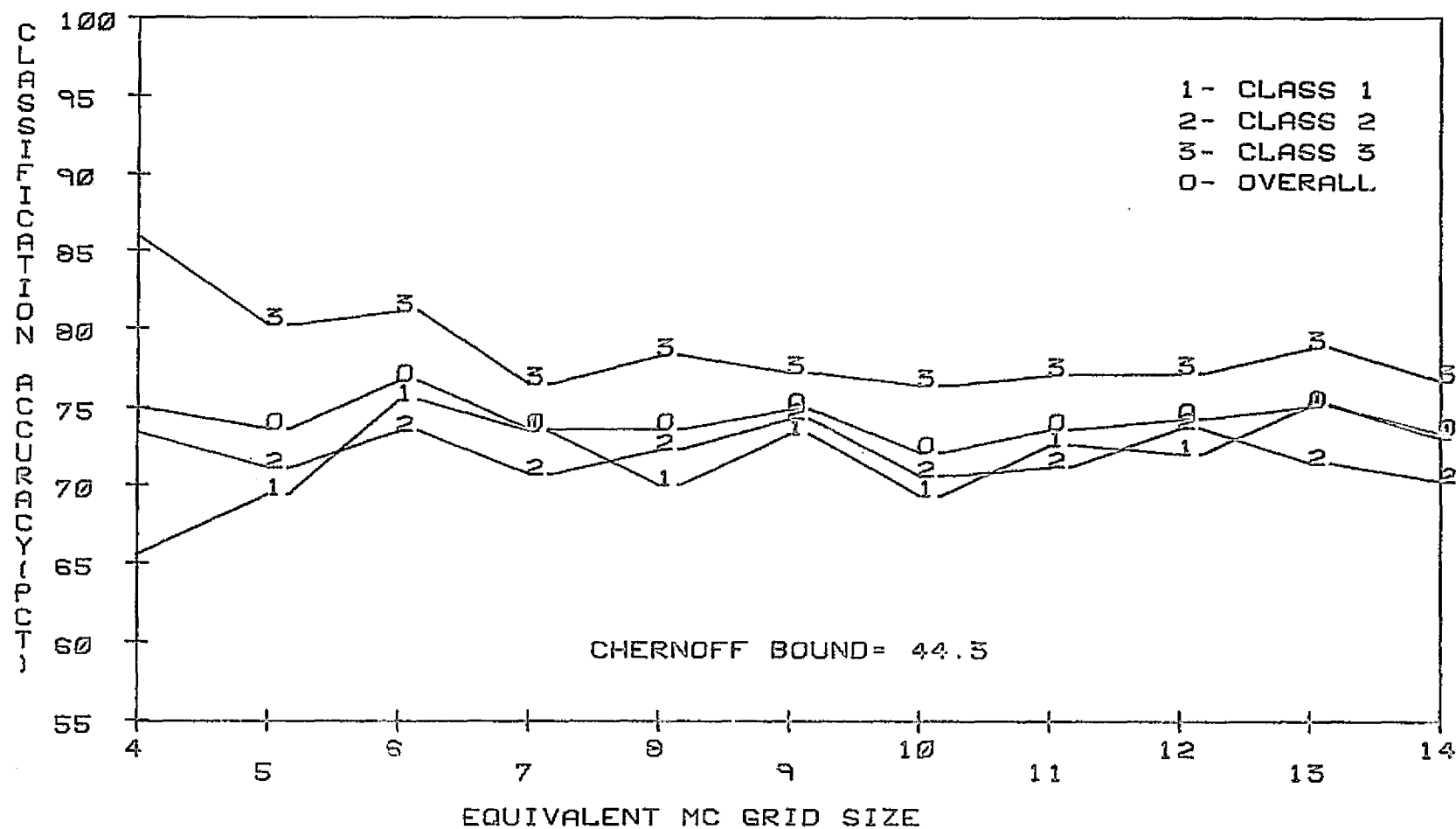


FIG. 4-18 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 5

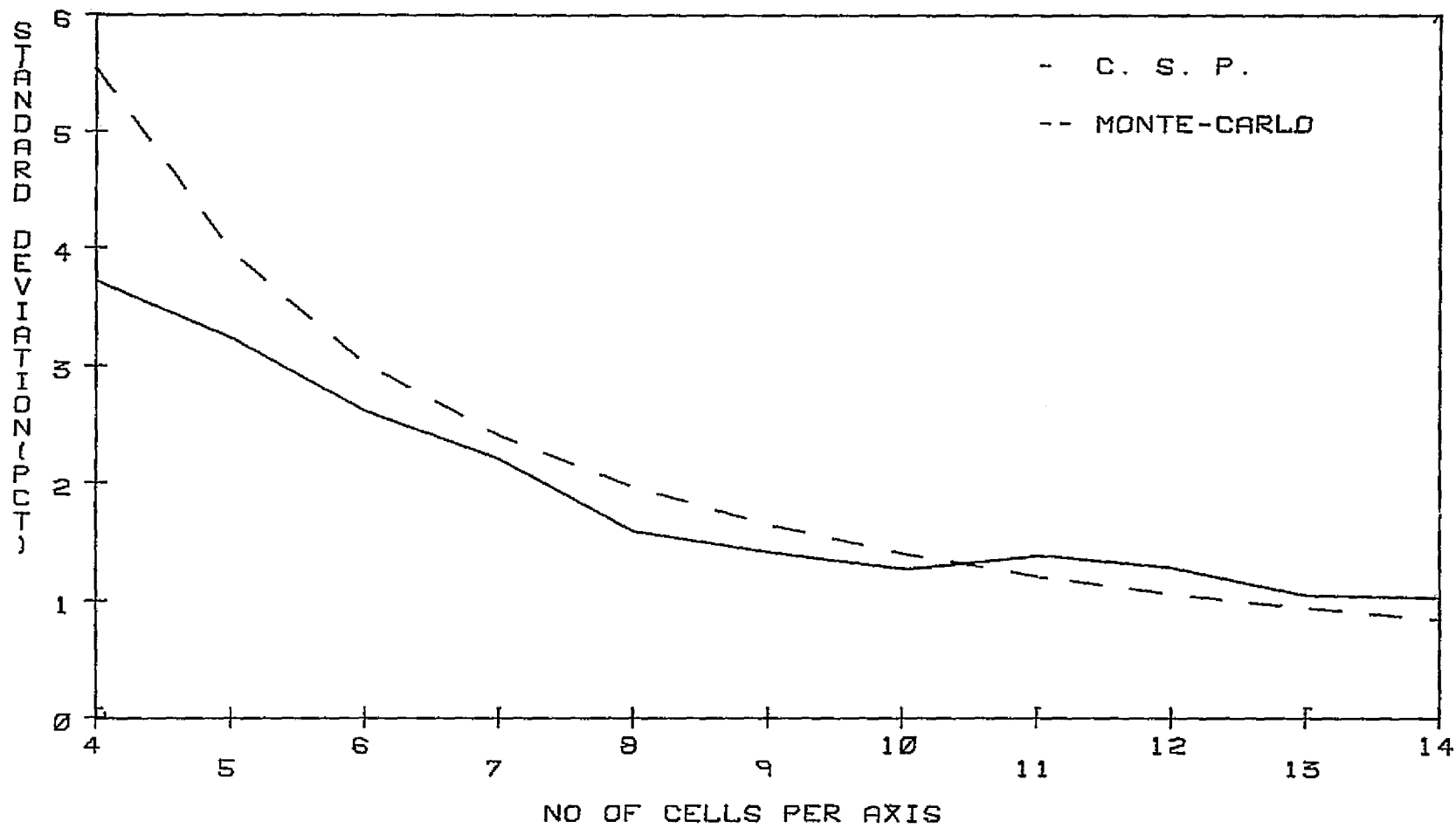


FIG. 4-19 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 5

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 4-12 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 6

G_S	$\hat{P}_{C \omega_1}$		$\hat{P}_{C \omega_2}$		$\hat{P}_{C \omega_3}$		$\hat{P}_C$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
4	59.9	67.2	64.3	71.9	73.2	89.1	65.8	76.0
5	67.6	64.5	71.5	71.1	77.1	80.2	72.1	71.9
6	65.6	69.4	69.9	73.5	75.5	82.1	70.3	75.0
7	66.5	67.9	70.6	72.5	77.6	79.3	71.6	73.3
8	65.6	64.3	68.4	72.3	77.3	79.5	70.4	77.0
9	65.9	65.6	69.4	75.3	77.4	78.3	70.9	73.1
10	66.5	65.3	70.2	71.1	79.0	77.9	71.9	71.5
11	67.8	67.9	70.5	71.9	79.7	78.2	72.7	72.7
12	66.7	67.4	72.2	73.7	79.7	78.8	72.8	73.3
13	66.6	70.2	72.0	72.0	79.1	80.6	72.5	74.3
14	67.8	67.9	72.6	71.2	78.3	78.9	72.9	72.7

TABLE 4-13 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

G_S	CSP	MC	$N_B\%$
4	2.5	5.9	41.2
5	2.1	4.2	29.2
6	1.9	3.2	31.8
7	1.5	2.5	23.5
8	1.4	2.1	23.0
9	1.0	1.7	19.1
10	1.2	1.5	19.3
11	0.8	1.3	16.3
12	0.8	1.1	15.5
13	0.7	1.0	13.6
14	0.6	0.9	13.1

ORIGINAL PAGE IS
OF POOR QUALITY

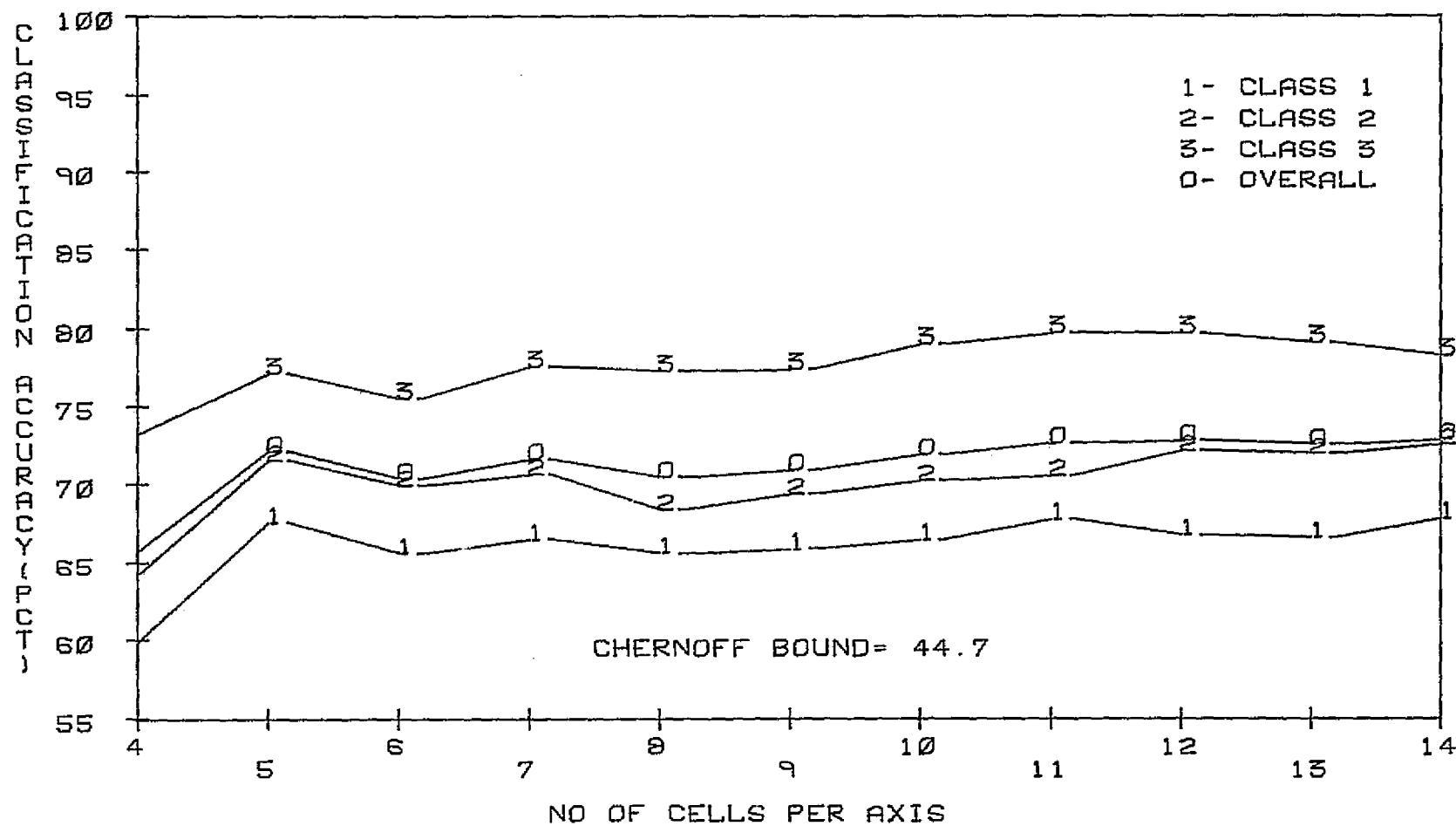


FIG. 4-20 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 6

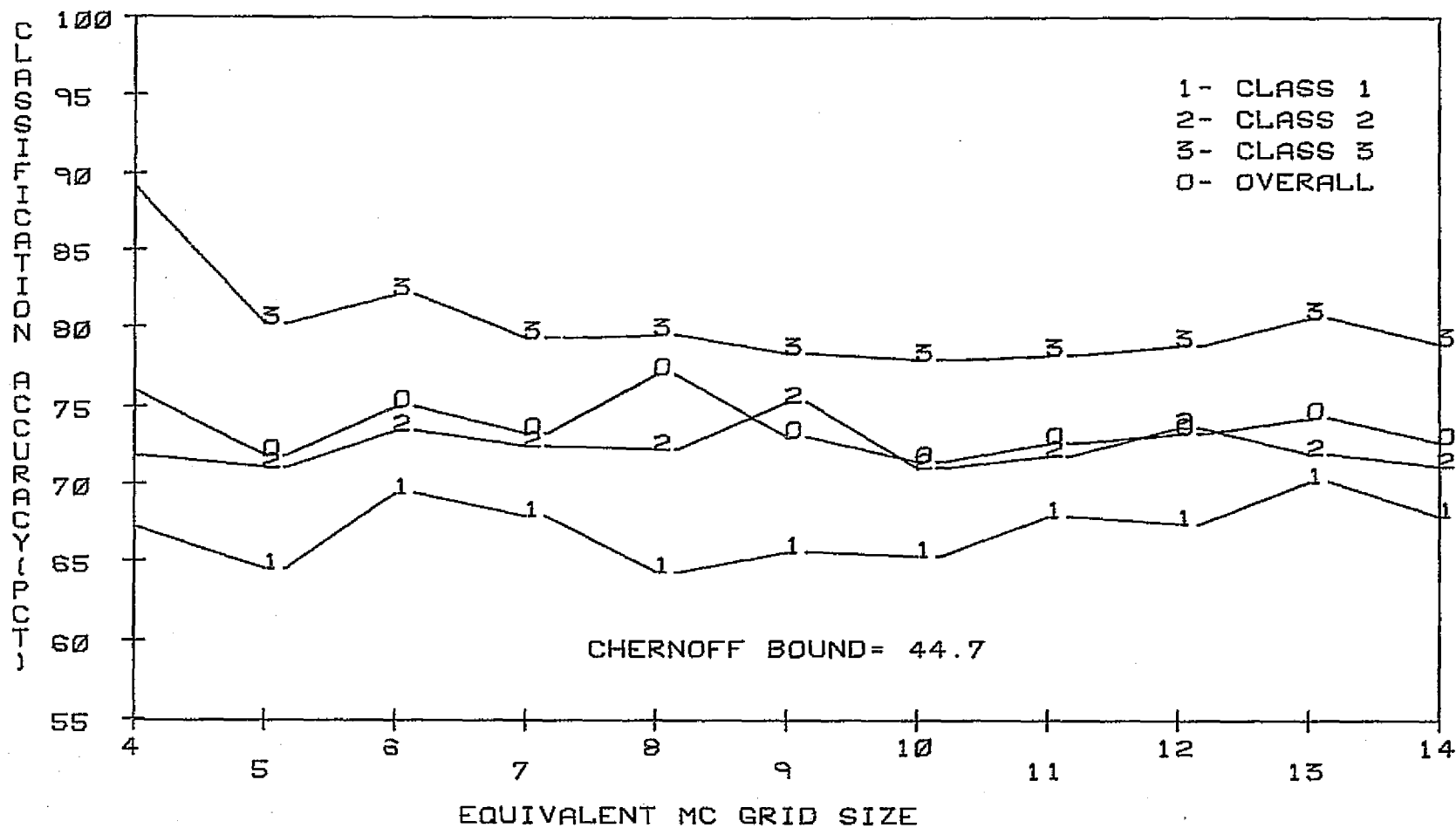


FIG. 4-21 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 6

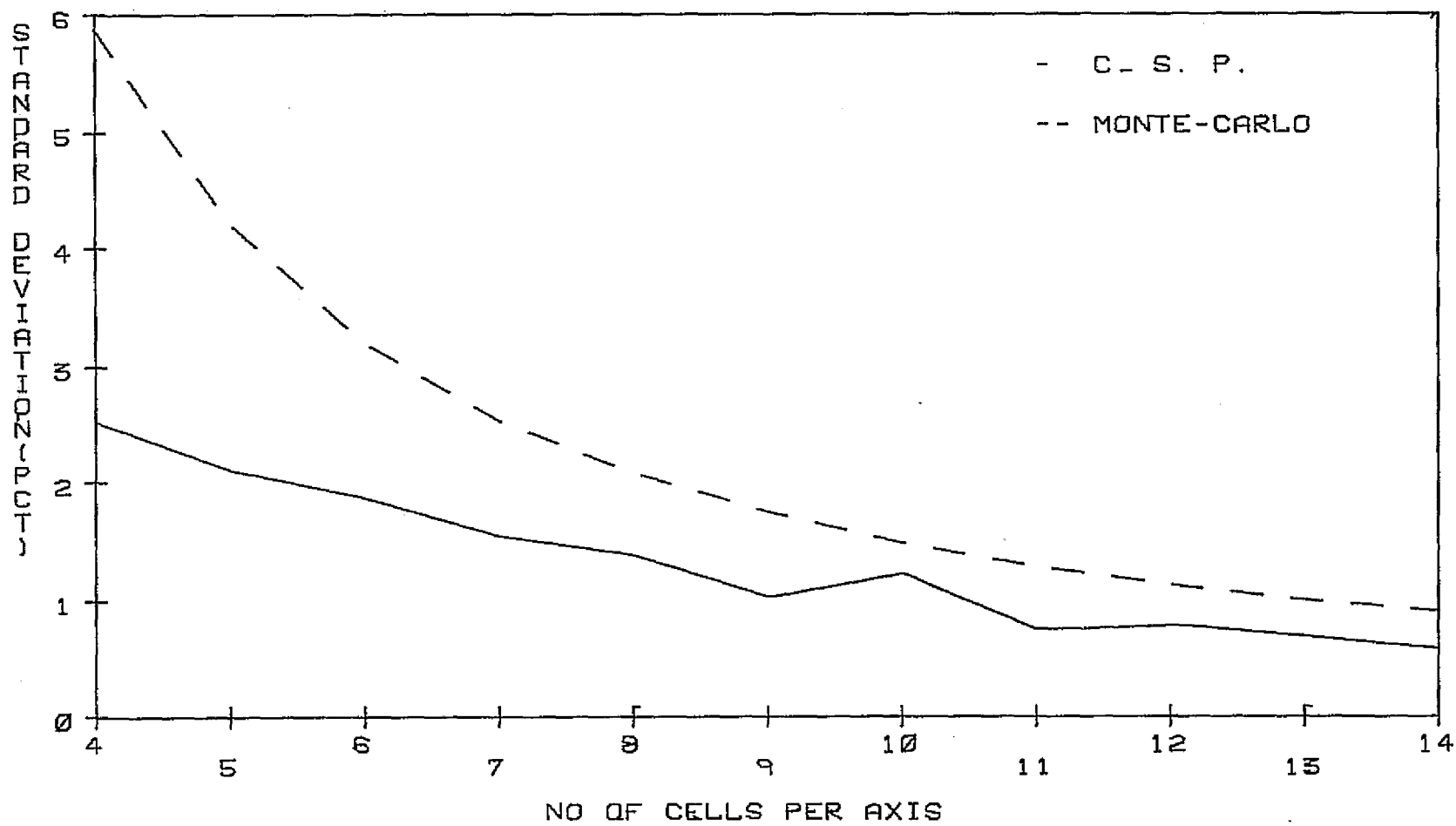


FIG. 4-22 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 6

TABLE 4-14 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 7

G_s	$\hat{P}_{C w_1}$		$\hat{P}_{C w_2}$		$\hat{P}_{C w_3}$		$\hat{P}_C$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
4	87.1	98.4	69.6	81.3	77.1	87.5	77.9	89.1
5	95.8	99.9	72.7	71.9	81.4	79.3	83.3	83.7
6	95.5	98.5	73.1	73.5	77.7	80.6	82.1	84.2
7	97.8	98.5	73.3	75.3	77.4	79.3	82.8	84.4
8	97.7	98.3	71.2	74.2	78.6	77.3	82.5	83.3
9	98.8	97.7	72.4	77.0	81.1	79.4	84.1	84.7
10	98.6	98.6	73.6	74.6	81.2	77.5	84.4	83.6
11	99.0	98.2	73.7	74.4	82.2	78.5	84.9	83.7
12	98.8	98.3	75.6	75.5	80.9	79.7	85.1	84.5
13	98.9	98.4	75.1	74.9	81.0	79.6	85.0	84.3
14	98.9	98.3	75.2	74.7	80.3	79.3	84.8	84.1

TABLE 4-15 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

G_s	CSP	MC	$N_B^{\%}$
4	3.3	1.2	27.0
5	2.2	0.9	23.5
6	1.2	0.7	23.0
7	1.4	0.5	19.1
8	1.4	0.4	18.5
9	0.9	0.4	16.7
10	0.9	0.3	15.9
11	0.8	0.3	14.7
12	0.5	0.2	13.9
13	0.5	0.2	13.2
14	0.5	0.2	12.5

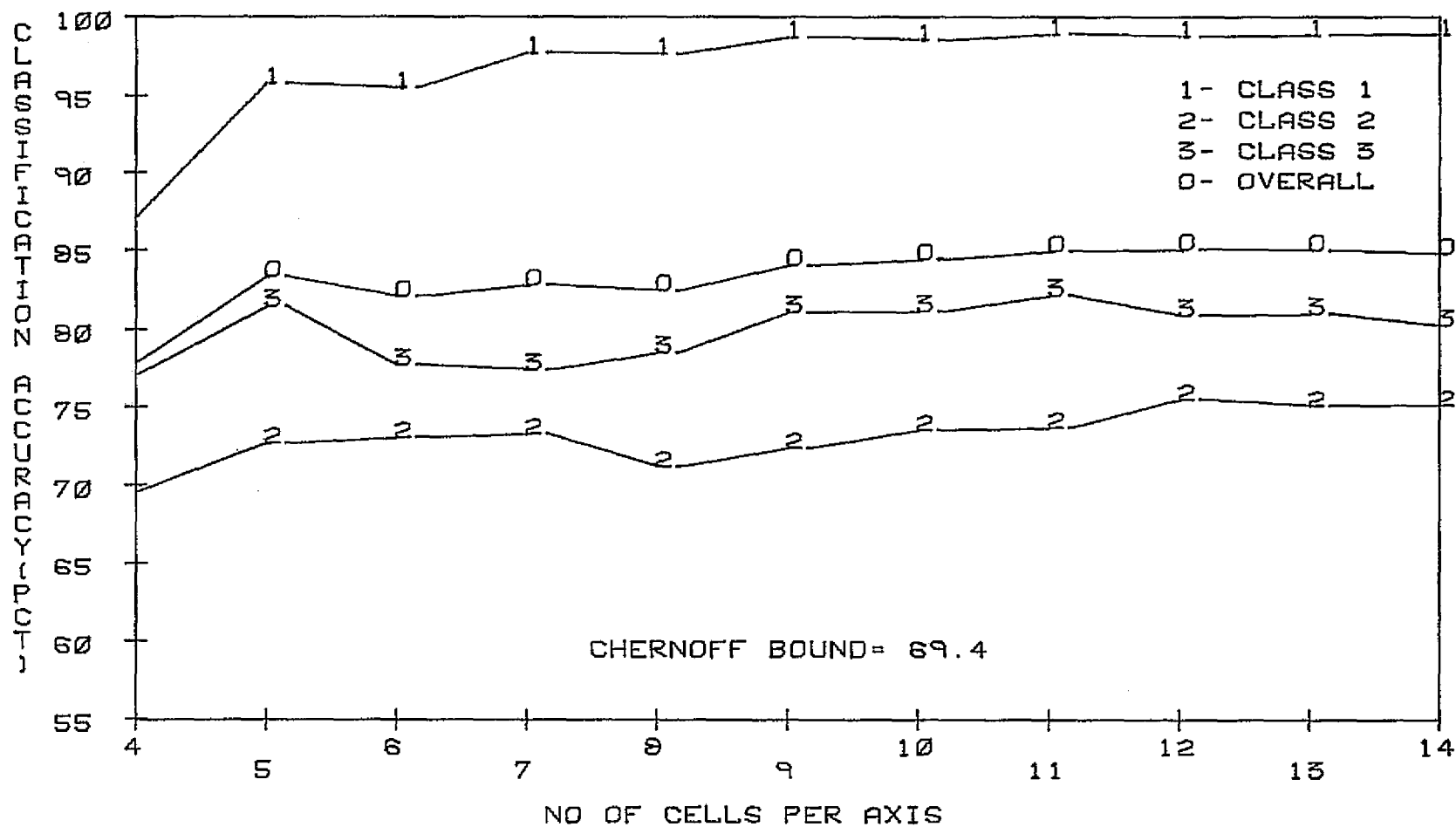


FIG. 4-23 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 7

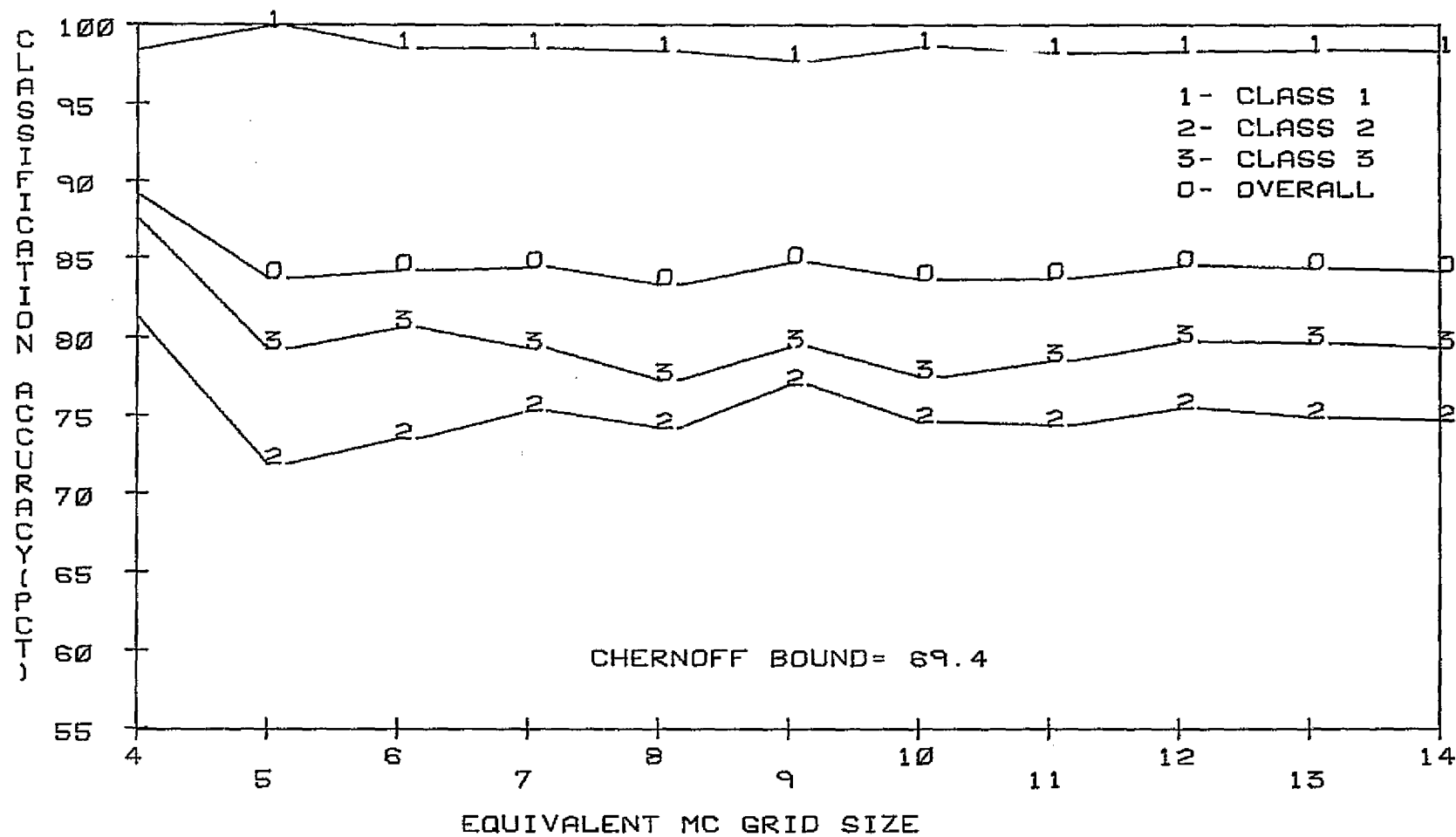


FIG. 4-24 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 7

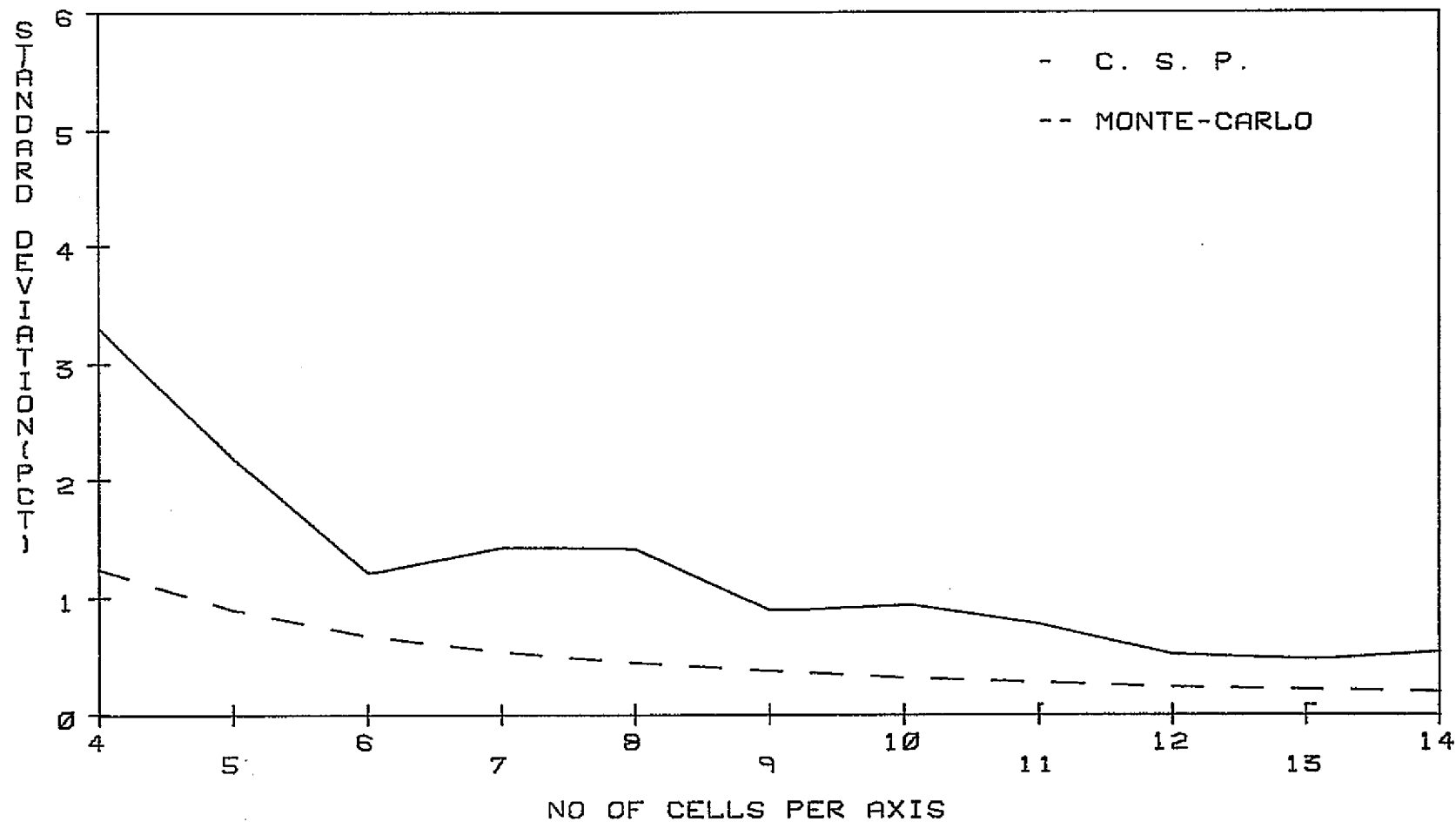


FIG. 4-25 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 7

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 4-16 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 8

G_s	$\hat{p}_{c \omega_1}$		$\hat{p}_{c \omega_2}$		$\hat{p}_{c \omega_3}$		$\hat{p}_c$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
4	86.0	92.2	69.6	81.3	68.5	89.1	74.7	87.5
5	95.6	97.5	74.1	87.2	75.1	70.7	81.6	81.0
6	94.2	98.0	73.4	76.0	74.6	79.6	80.7	84.5
7	97.5	97.5	74.4	76.9	79.2	76.9	83.7	83.7
8	97.4	96.3	72.5	77.1	78.8	73.6	82.9	82.3
9	97.7	97.3	73.3	79.1	80.0	77.1	83.7	84.5
10	97.8	98.1	75.0	76.7	80.5	74.8	84.4	83.2
11	98.4	97.5	75.7	76.2	78.4	76.5	84.2	83.4
12	98.3	97.5	76.6	78.4	77.7	78.2	84.2	84.7
13	98.1	96.8	76.6	76.6	78.3	78.7	84.3	84.0
14	98.3	97.0	76.9	77.1	78.1	75.9	84.4	83.3

TABLE 4-17 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

G_s	CSP	MC	$N_B\%$
4	2.3	1.8	41.8
5	2.4	1.3	24.8
6	1.8	0.9	28.6
7	1.5	0.8	22.2
8	1.2	0.6	22.4
9	1.0	0.5	18.4
10	0.8	0.4	19.1
11	0.8	0.4	15.9
12	0.9	0.3	19.0
13	0.5	0.3	14.9
14	0.3	0.3	17.5

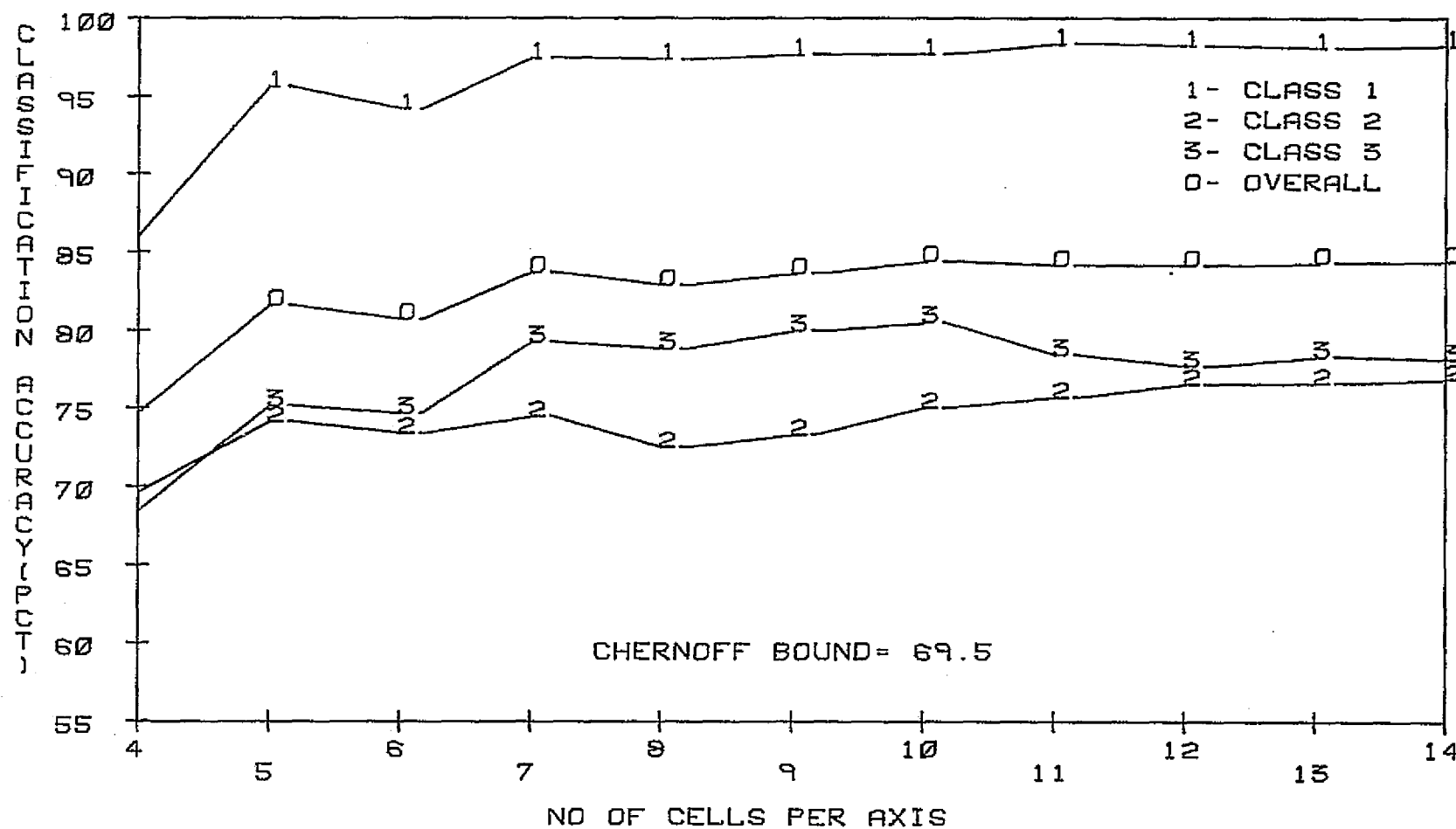


FIG. 4-26 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 9

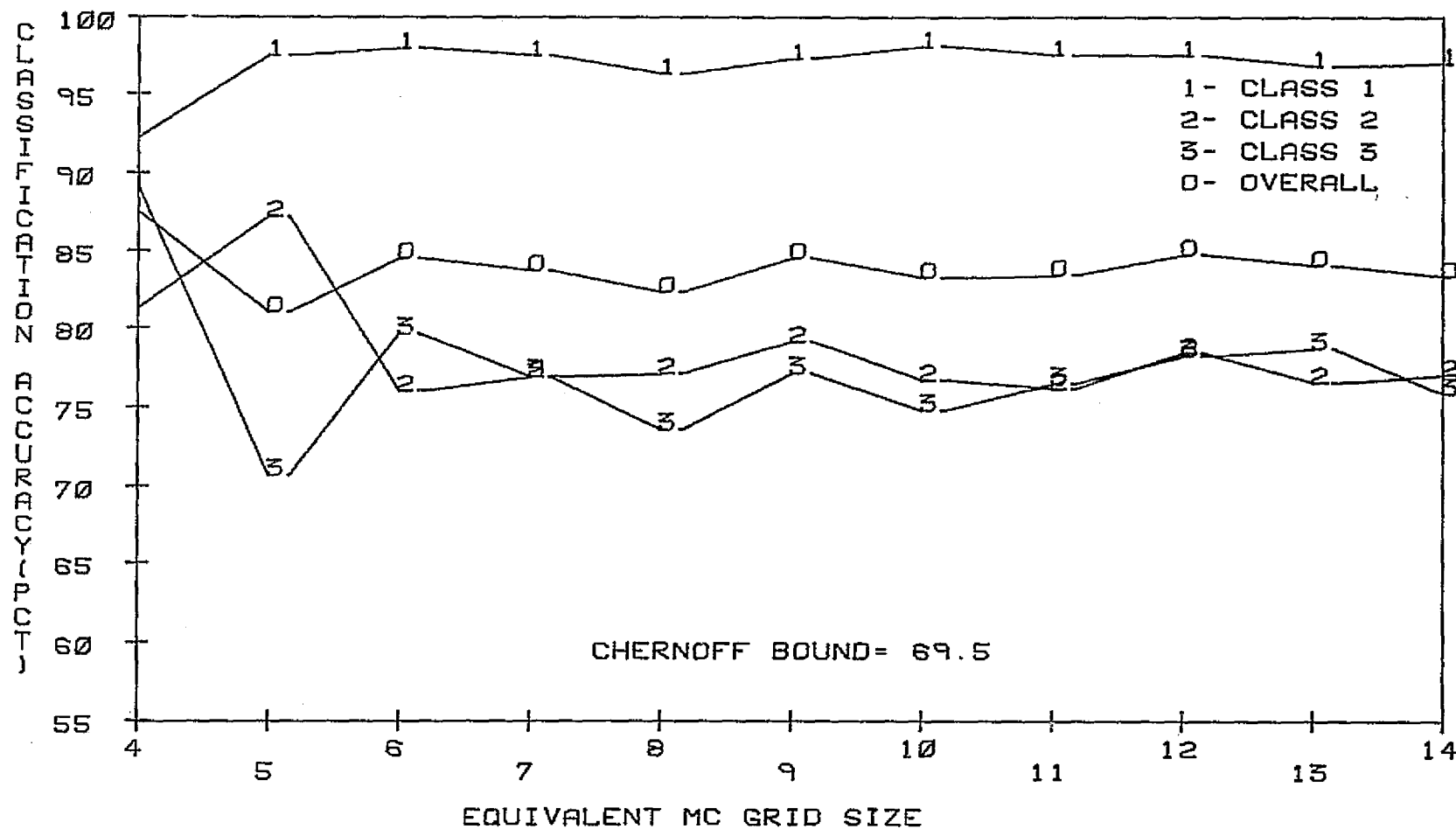


FIG. 4-27 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 8

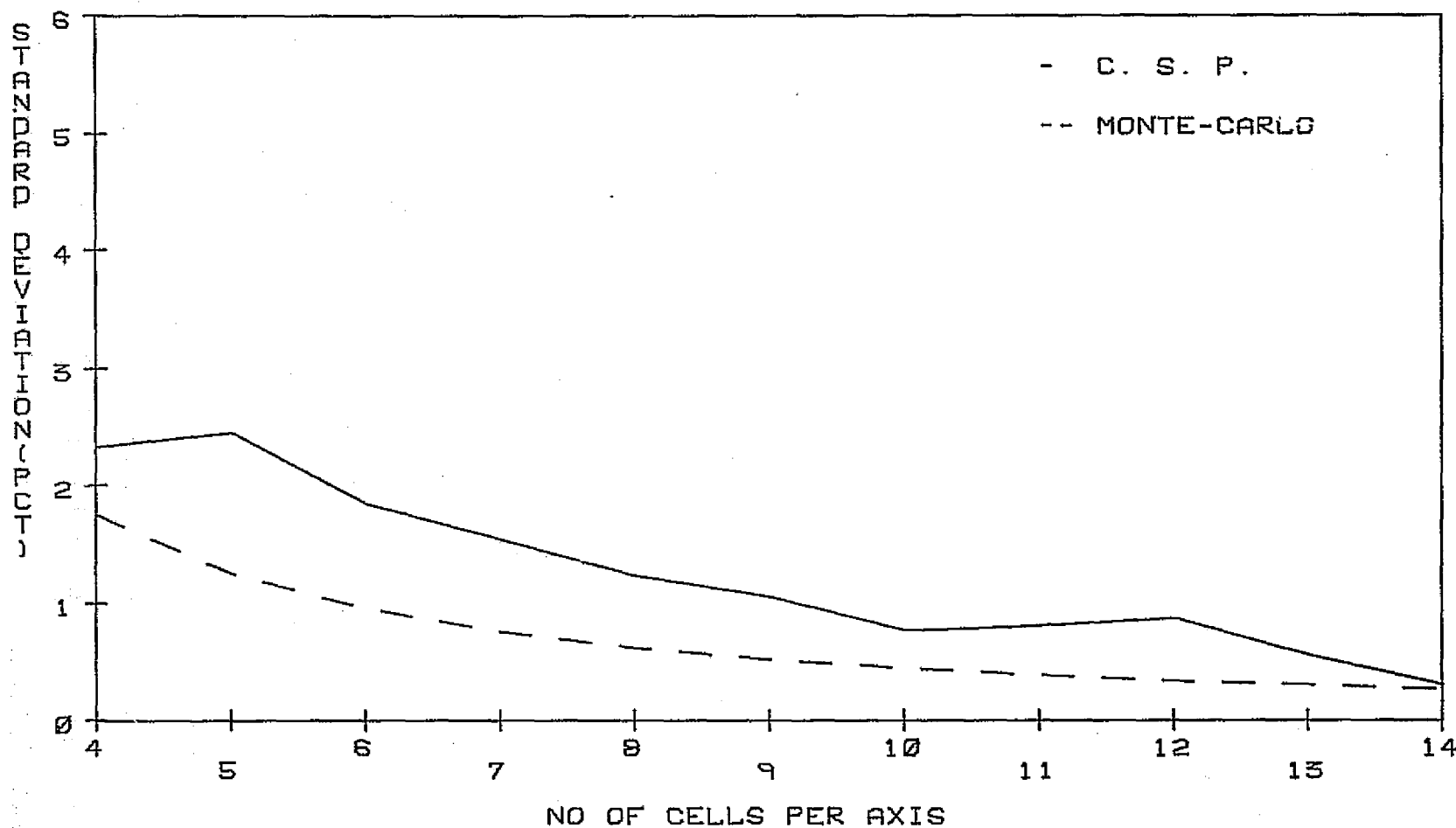


FIG. 4-28 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 8

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 4-18 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 9

$\hat{p}_{c w_1}$	$\hat{p}_{c w_2}$		$\hat{p}_{c w_3}$		$\hat{p}_c$	
$\frac{G}{S}$	CSP	MC	CSP	MC	CSP	MC
4	87.4	99.9	73.4	76.0	92.2	77.6
5	96.1	99.2	74.0	71.9	81.8	82.9
6	95.6	99.5	69.4	74.0	83.2	81.6
7	98.3	99.4	71.7	74.1	82.4	83.5
8	98.2	99.6	70.6	73.6	81.3	83.4
9	99.0	98.8	72.5	77.0	82.4	84.7
10	98.9	99.2	72.6	73.3	83.8	85.1
11	99.3	99.2	73.3	73.5	83.1	85.2
12	99.2	99.2	74.1	75.7	82.4	85.2
13	99.4	99.0	74.5	73.7	82.5	85.5
14	99.3	98.8	74.7	74.5	82.7	85.6

TABLE 4-19 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

$\frac{G}{S}$	CSP	MC	N_B
4	2.5	1.0	31.1
5	2.3	0.7	21.1
6	1.9	0.6	24.0
7	1.5	0.4	18.2
8	1.4	0.4	18.2
9	0.9	0.3	19.8
10	0.9	0.3	15.7
11	0.6	0.2	16.5
12	0.8	0.2	13.8
13	0.5	0.2	14.3
14	0.4	0.2	12.4

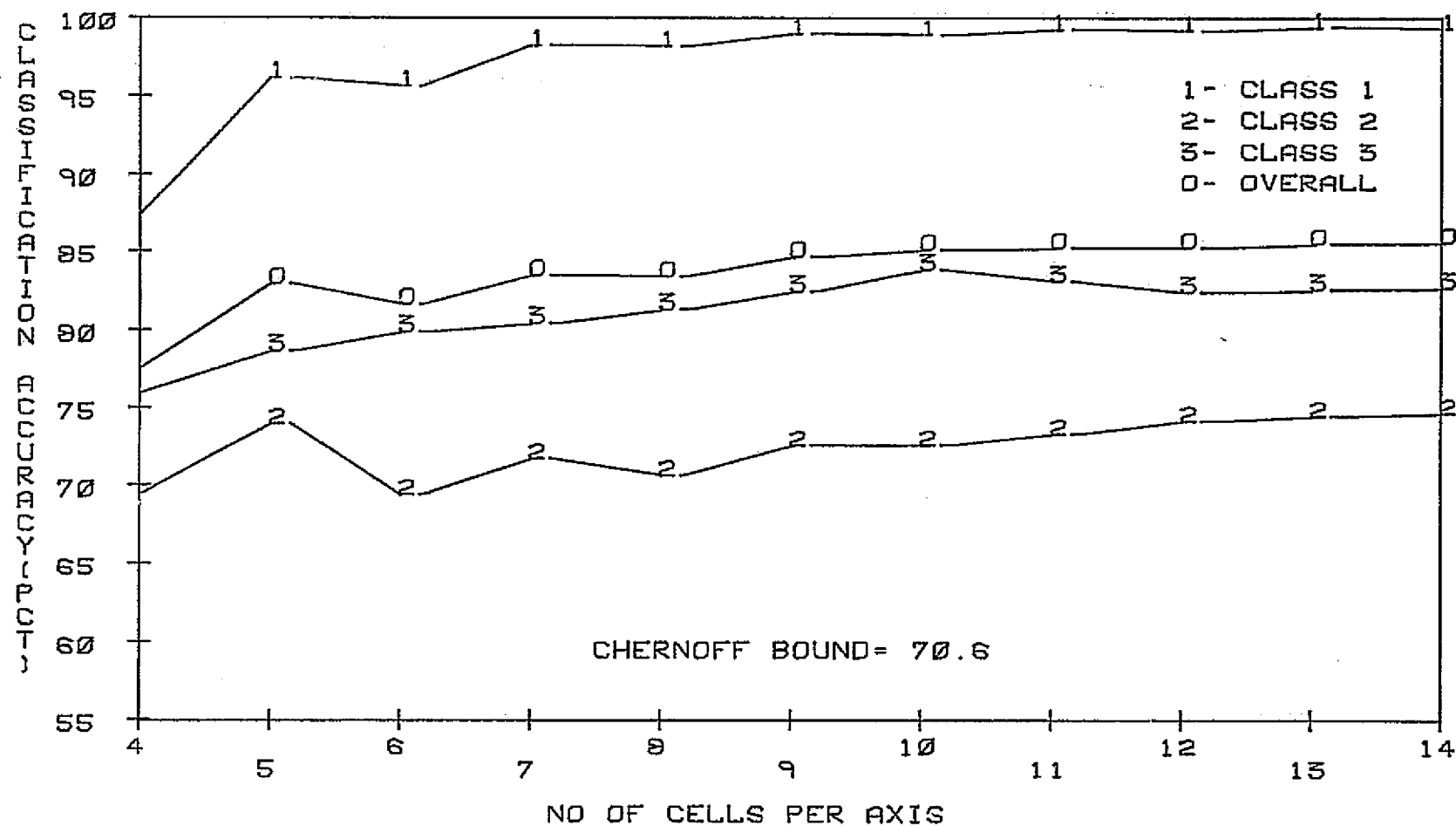


FIG. 4-29 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 9

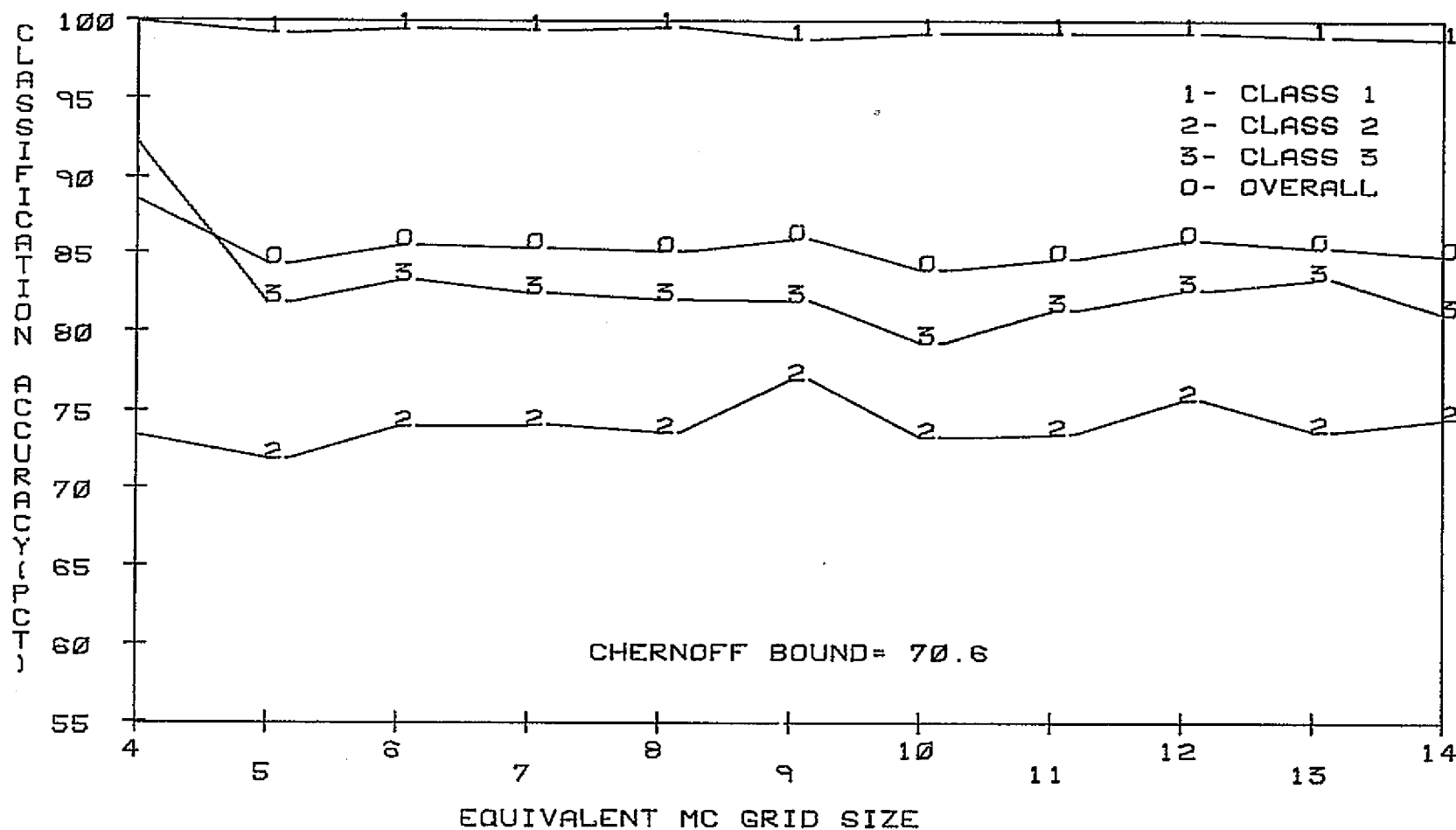


FIG. 4-30 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 9

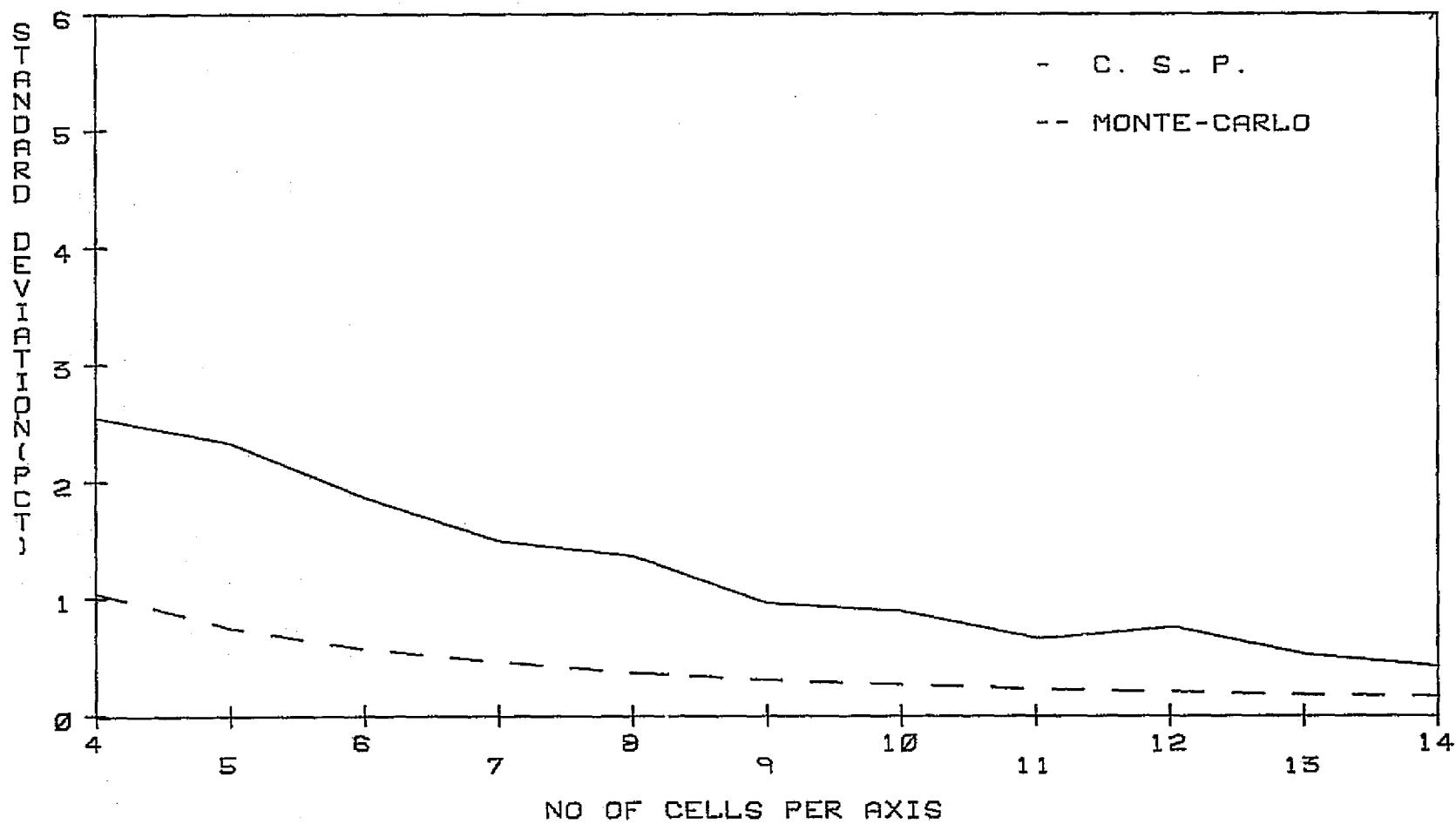


FIG. 4-31 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 9

TABLE 4-20 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 10

C_S	$\hat{P}_C w_1$		$\hat{P}_C w_2$		$\hat{P}_C w_3$		$\hat{P}_C$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
4	87.0	99.9	69.6	75.0	73.5	90.6	76.7	88.5
5	96.3	98.3	73.7	71.1	80.9	81.0	83.6	83.5
6	95.7	99.5	73.3	75.5	80.7	83.2	83.2	86.1
7	97.7	99.9	74.6	72.2	81.7	80.2	84.7	84.2
8	98.0	99.4	72.3	71.7	82.9	80.8	84.4	84.0
9	99.1	98.6	72.5	75.9	81.9	81.2	84.5	85.2
10	99.0	98.9	74.0	73.7	81.8	81.0	84.9	84.5
11	99.0	99.2	74.7	73.2	80.8	79.9	84.9	84.1
12	99.0	98.5	75.1	75.1	81.3	82.4	85.1	85.3
13	99.2	98.7	74.7	72.6	81.8	83.3	85.2	84.9
14	99.3	98.7	74.6	73.9	82.4	81.6	85.4	84.7

TABLE 4-21 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

G_S	CSP	MC	$N_B\%$
4	2.5	1.0	33.3
5	2.3	0.7	20.0
6	1.8	0.6	22.3
7	1.3	0.4	18.2
8	1.4	0.4	18.9
9	1.0	0.3	15.0
10	0.8	0.3	15.5
11	0.6	0.2	13.2
12	0.6	0.2	13.3
13	0.6	0.2	11.6
14	0.6	0.2	11.5

ORIGINAL PAGE IS
OF POOR QUALITY

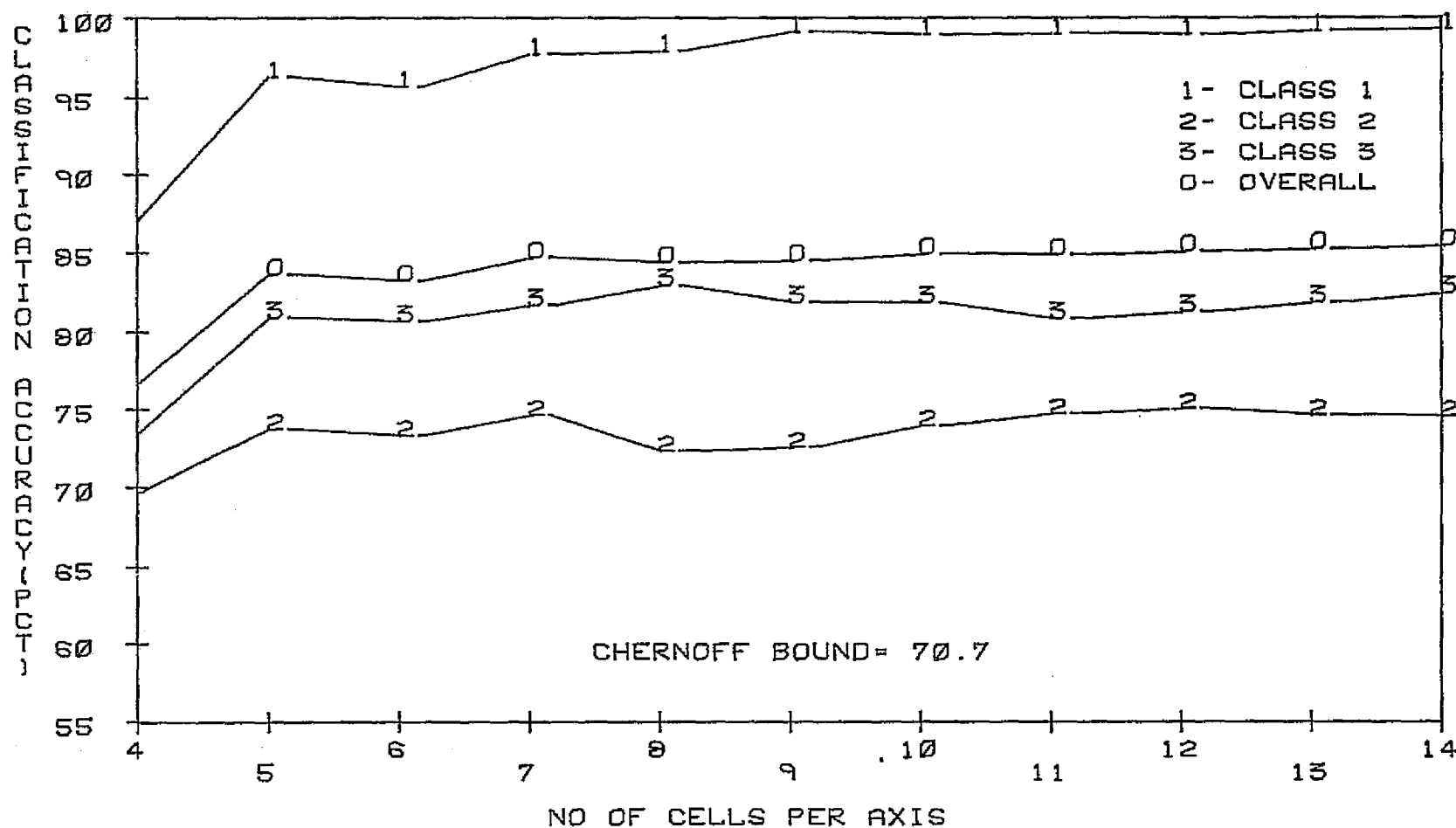


FIG. 4-32 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 10

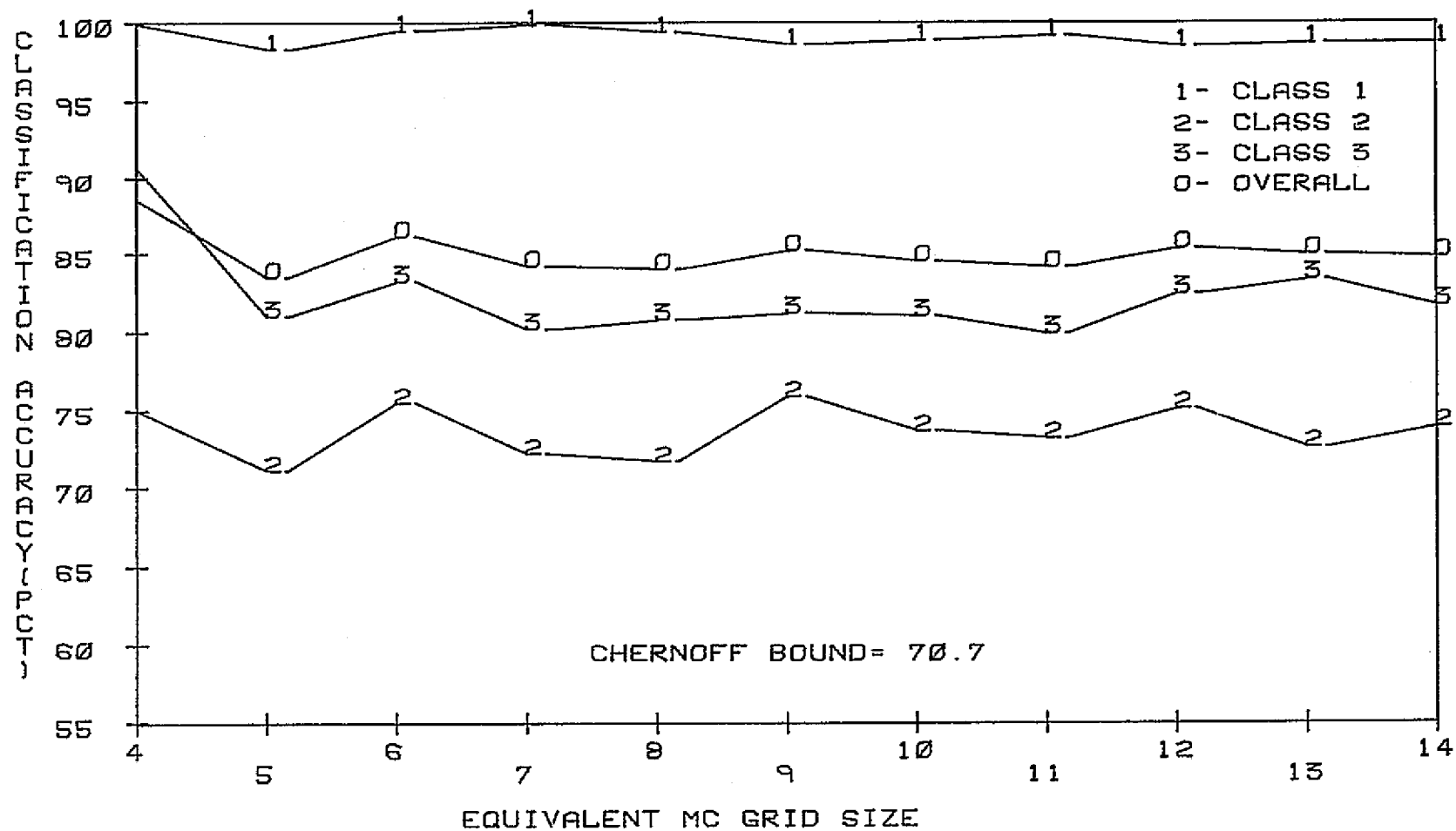


FIG. 4-33 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 10

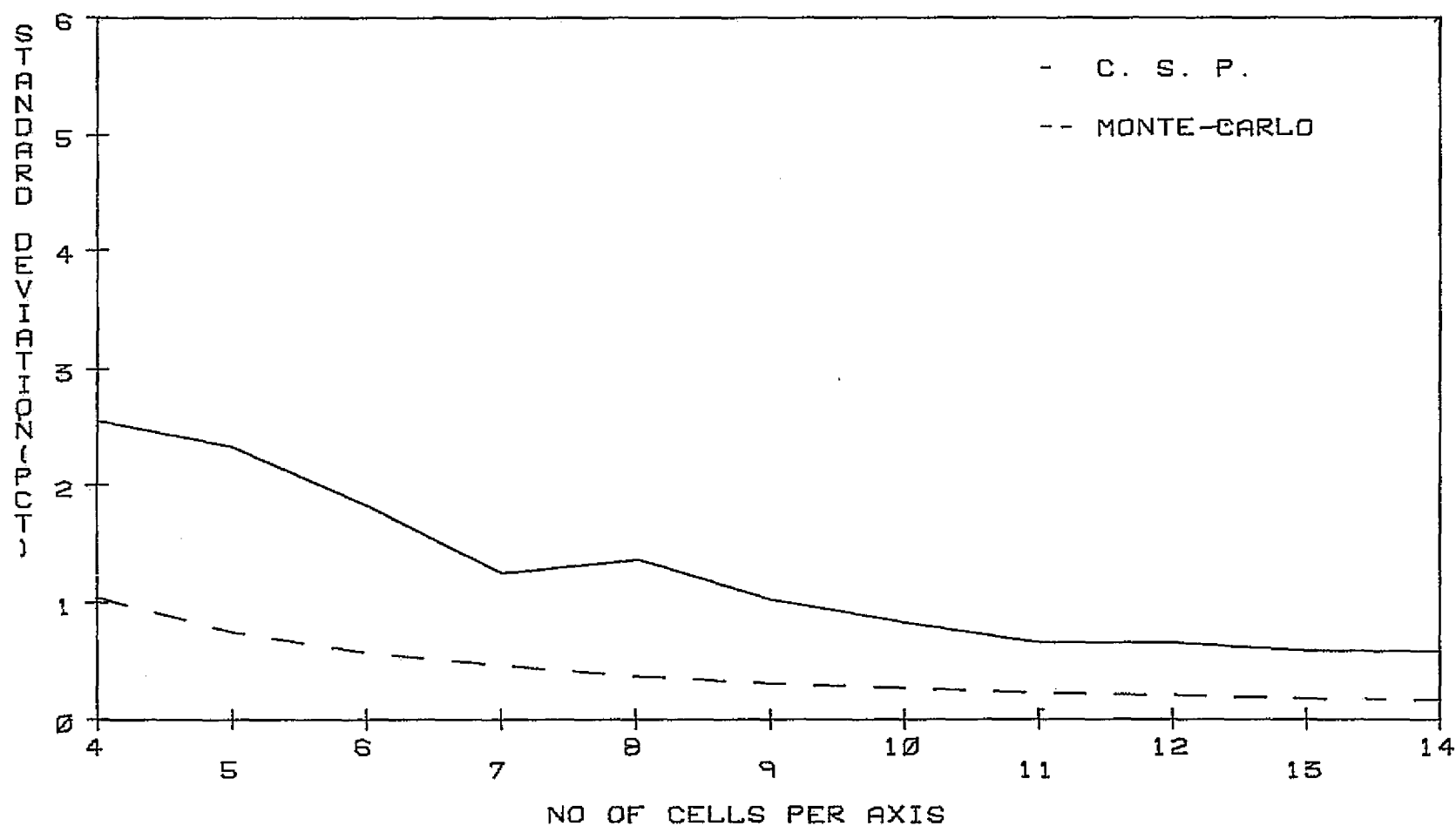


FIG. 4-34 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 10

TABLE 4-22 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP AND MC ESTIMATION TECHNIQUES. CASE 11

G_S	$\hat{P}_C \omega_1$		$\hat{P}_C \omega_2$		$\hat{P}_C \omega_3$		$\hat{P}_C$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
4	87.5	98.4	69.4	79.7	78.5	92.2	78.5	90.1
5	96.3	99.2	73.7	74.4	80.9	83.5	83.6	85.7
6	95.9	99.5	71.6	76.0	80.5	85.2	82.7	86.9
7	98.7	99.9	73.0	74.7	81.5	82.4	84.4	85.7
8	98.6	99.8	72.2	73.8	82.7	83.5	84.5	85.7
9	99.4	99.7	72.4	77.9	83.3	84.2	85.0	87.3
10	99.4	99.9	73.6	75.1	84.5	82.9	85.8	86.0
11	99.6	99.5	74.3	74.1	84.2	82.5	86.1	85.4
12	99.6	99.8	75.6	76.4	84.0	84.7	86.4	87.0
13	99.7	99.7	75.7	75.2	83.9	84.9	86.5	86.6
14	99.7	99.7	75.9	75.1	83.8	83.4	86.5	86.1

TABLE 4-23 PERCENT CSP AND MC STANDARD DEVIATIONS ACHIEVED FOR CLASS 1.

G_S	CSP	MC	$N_B \%$
4	2.3	0.7	25.0
5	2.3	0.5	20.0
6	1.8	0.4	18.8
7	1.6	0.3	16.1
8	1.4	0.2	16.6
9	1.2	0.2	14.8
10	1.3	0.2	16.8
11	0.9	0.1	15.2
12	0.7	0.1	16.0
13	0.3	0.1	15.1
14	0.7	0.1	15.3

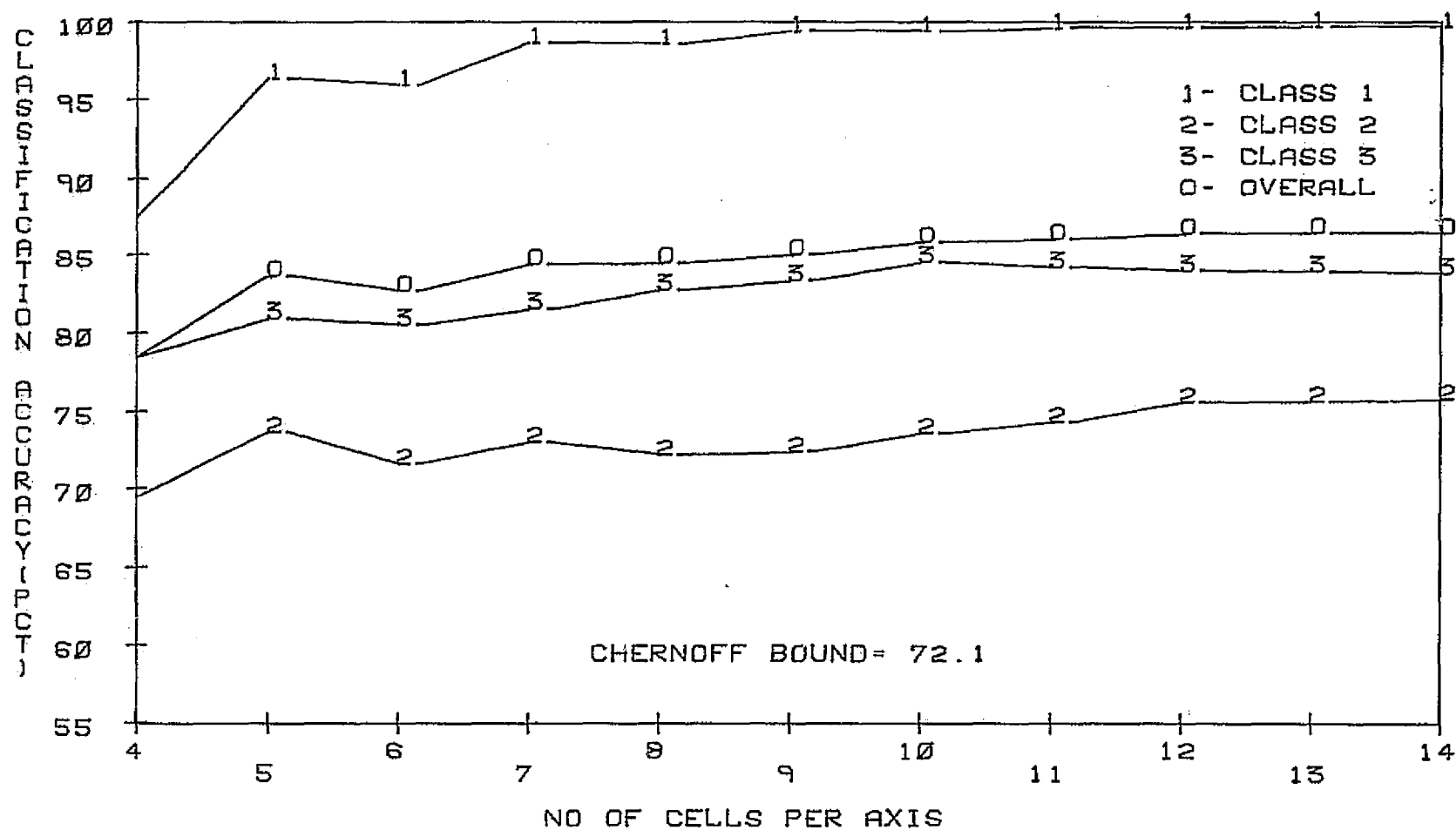


FIG. 4-35 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE SCATTER. CASE 11

ORIGINAL PAGE IS
OF POOR QUALITY

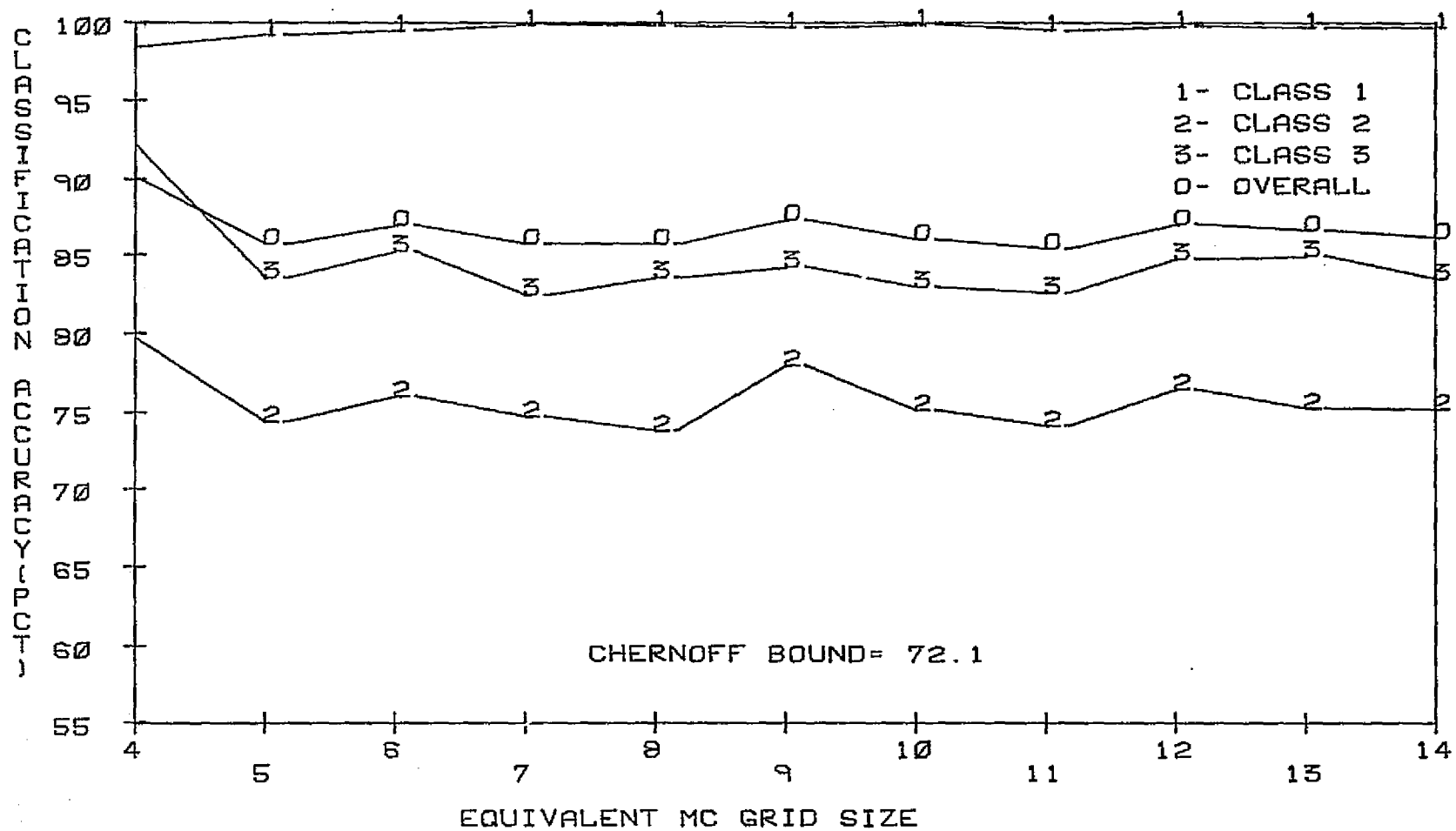


FIG. 4-36 MC CLASSIFICATION ACCURACY ESTIMATE VS. SAMPLE SIZE.
VARIABLE SCATTER. CASE 11

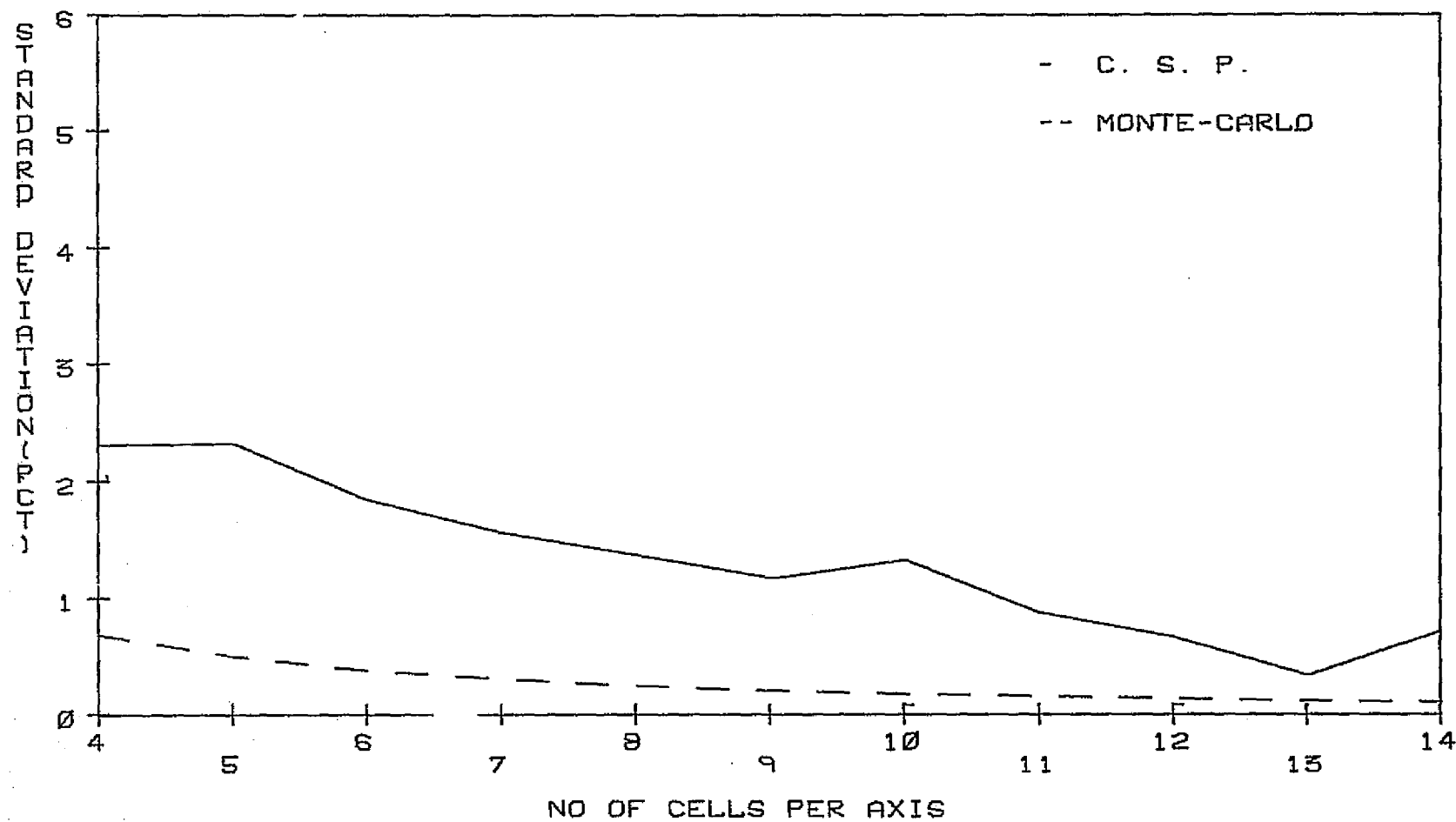


FIG. 4-37 MC AND CSP ERROR ESTIMATE STANDARD DEVIATIONS.
VARIABLE SCATTER. CASE 11

TABLE 4-24 COMPARISON OF CSP AND MC PERCENT CLASSIFICATION ACCURACY.
VARIABLE SCATTER.

Case	$\hat{P}_c \omega_1$		$\hat{P}_c \omega_2$		$\hat{P}_c \omega_3$		$\hat{P}_c$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
1	68.6	69.3	74.3	74.1	75.9	76.7	72.9	73.4
2	64.5	63.6	73.4	74.1	77.8	78.5	71.9	72.1
3	78.0	75.8	71.7	70.9	76.3	75.9	75.3	74.2
4	66.1	64.4	73.3	73.2	79.1	79.1	72.8	72.2
5	73.0	73.0	71.7	70.2	77.1	76.7	73.8	73.3
6	67.8	67.9	72.6	71.2	78.3	78.9	72.9	72.7
7	98.9	98.3	75.2	74.7	80.3	79.3	84.8	84.1
8	98.3	97.0	76.9	77.1	78.1	75.9	84.4	83.3
9	99.3	98.8	.	74.5	82.7	81.1	85.6	84.8
10	99.3	98.7	74.6	73.9	82.4	81.6	85.4	84.7
11	99.7	99.7	75.9	75.1	83.8	83.4	86.5	86.1

that in any MC simulation process such small sample sizes are not used anyway, but it is precisely the reasons outlined above that makes employment of a large data base mandatory. This requirement would not be overly restrictive with unlimited computation time. Since this is generally not the case, adequate performance with small sample sizes becomes a significant property. The sufficiency of small grid sizes for adequate performance was expected considering the structure of the sampling grid. In sec. 2.5 this matter was discussed and it was pointed out that the grid is a partitioned hypercube with each edge $2\sqrt{n}\sigma_i$ long. Therefore, small values of n are capable of sampling substantial portions of the feature space.

The next systematic feature in the variation of the CSP estimator is a periodic oscillations for each increment of the grid size. This phenomenon, like most other properties of this estimator, is the product of the geometry of the problem. As described before, the rule governing the assignment of sampling cells to a particular domain can potentially exclude (include) the entire cell even though only a portion of it lies outside (inside). The grid, being a dynamic structure, interacts with the fixed boundaries to produce the oscillatory character of the estimate the manner and intensity of which depends on the shape of the boundaries involved. Among all the cells that are located around the contours of Γ_i there are always some excluded from the

inside domain but the centers of which are close enough to the boundary such that one increment in the grid size would move them to the inside. This outside-to-inside shift would turn an underestimating grid to an overestimating one. The size of this step depends on the number of cells capable of making this shift. It is not hard to see that with a grid composed of elements with linear features, the worst case occurs when the boundaries themselves are linearly structured. In fact, when the feature space is divided by a set of hyperplanes, this periodic cycle can take on substantial amplitude and hence provide a worst case situation for this algorithm. This is in general a very minor limitation due to the fact that actual remotely sensed data, and most data in general where the information itself is a realization of a stochastic process, are unlikely to be optimally classified into region bounded by hyperplanes. Estimation of the numerical values of the estimates for various cases here shows that after a steady state value has been reached, the magnitude of the oscillation peaks are well within 1 percent.

The variation of the MC estimate with the sample size exhibits less recognizable features mainly due to the underlying randomness of the process. What is particularly different is the absence of the initial rise of the classification accuracy estimate. This observation should be viewed with caution, however, due to the small sample

sizes involved. In order to use the results of these estimates in a conclusive manner, any comparisons made should be restricted to samples of greater than 1000 (equivalent to 10 cells per axis) which in that range the estimate exhibits an adequately small variance.

One topic yet unexplored is how close is the CSP estimate of the classification accuracy to the Bayes estimate. In the general case under study, the availability of such reference is quite limited and in fact count estimators are the only alternative. Therefore, the availability of the MC estimation results makes the required comparison feasible. Table 4-24 lists 4 classification accuracy estimates obtained via CSP (MC) techniques for the highest grid size (sample size). Throughout this table, the values of the two estimates are quite close and in two cases ($\hat{P}_c|_{\omega_1}$ Case 5 and $\hat{P}_c|_{\omega_2}$ Case 4) the results are identical to one significant figure. The differential for $\hat{P}_c$ ranges from a low of 0.2% for Case 2 and 6 to a high of 1.1% for Case 3 and 8. Averaged over all the cases, this difference amounts to 0.6%.

One of the most desirable properties of any estimator is consistency. The error variance is calculated using (2-37) and is plotted for all the 11 cases. Examination of these plots clearly shows that the variances of the CSP estimates are monotonically decreasing as the number of cells per axis increases. This is particularly

significant because as discussed in sec. 2.4.1 (2-37) does not conclusively indicate that $\lim_{n \rightarrow \infty} \text{Var}\{e_T\} \rightarrow 0$ although it strongly suggests that. This property is brought about by the fact that the total number of the points on the boundary, N_B , as a percentage of the points inside, monotonically decreases with increasing grid size. This observation is consistent with the assertion that the boundary cells are the only error causing elements in the CSP algorithm. Comparing the CSP and MC error variances for different cases, several properties are distinguished. The CSP error variance, for the medium recognition rates, is generally below that of the MC technique. The rate at which the MC variance falls, however, is faster and thus if their initial values are close a crossover takes place for large sample sizes. This difference in the rate of decay is evident from the expression for CSP error variance. Rewriting (2-37)

$$\text{Var}\{e_T\} = \frac{1}{12} \left(\frac{2}{n\pi}\right)^n \sum_{i=1}^{N_B} f_i^2 \quad (4-7)$$

The corresponding variance for a MC estimator is given by

$$\text{Var}\{e_T\} = \frac{\epsilon(1-\epsilon)}{N_t} \quad (4-8)$$

where ϵ is the Bayes probability of error. Noting that $N_t = n^N$, it is clear that both estimators fall off at a

rate of $1/N_t$. The CSP error variance, however, has another sample dependent term, $\sum f_i^2$, which steadily increases with increasing N_t and thus cancels $1/N_t$ term to some extent, hence slower convergence. In the high classification accuracy bracket both estimators have small variances with that of MC slightly below CSP. This property is due to the fact that $\epsilon(1-\epsilon)$ is the dominant factor for small N_t . Once the initial value of the MC error variance is smaller, its faster fall off would keep it below that of CSP. The differences involved, however, are small. Selecting a medium size grid, the absolute value of standard deviation differential ranges from a high of 1.16% for Case 11 to a low of 0.13% for Cases 3 and 5.

4.2.2 Fixed Scatter, Variable Mean

In order to observe the variation of the probability of misclassification with changes in the mean of a population, the simplex arrangement of Fig. 4-2 was maintained along with the fixed statistics of ω_2 and ω_3 . Case 1 of sec. 4.2.1, the smallest separability, was selected as an initial starting point and the nonzero component of μ_1 , m_1 , was incremented by 0.10 step each time. A total of 7 cases ranging from $\bar{J} = .55$ to $\bar{J} = 0.96$ were covered and are listed in Table 4-25. Similar to the variable scatter case, the classification error estimate is obtained using CSP and MC techniques. In order to avoid duplication

TABLE 4-25 TEST CASES ARRANGED BY INCREASING SEPERABILITY.
VARIABLE MEAN.

m_1	$\bar{J}$	$\bar{J}^*$	C_B^*
0.8	0.55	57	41.8
0.9	0.60	62	44.7
1.0	0.66	68	47.5
1.1	0.73	76	51.4
1.2	0.80	87	53.2
1.3	0.88	100	55.9
1.4	0.96	100	58.4

TABLE 4-26 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP
ESTIMATION TECHNIQUE. CASE 12

G_s	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{F}_c$
4	57.6	69.6	71.2	66.2
5	66.8	72.3	75.5	71.5
6	67.0	71.0	74.5	70.8
7	71.0	71.4	76.2	72.9
8	72.3	69.6	76.5	72.8
9	72.9	70.4	77.0	73.4
10	73.2	72.6	77.4	74.4
11	72.7	73.2	77.4	74.4
12	72.1	74.2	76.8	74.4
13	72.7	74.3	77.6	74.9
14	72.1	74.4	76.9	74.5

ORIGINAL PAGE IS
OF POOR QUALITY

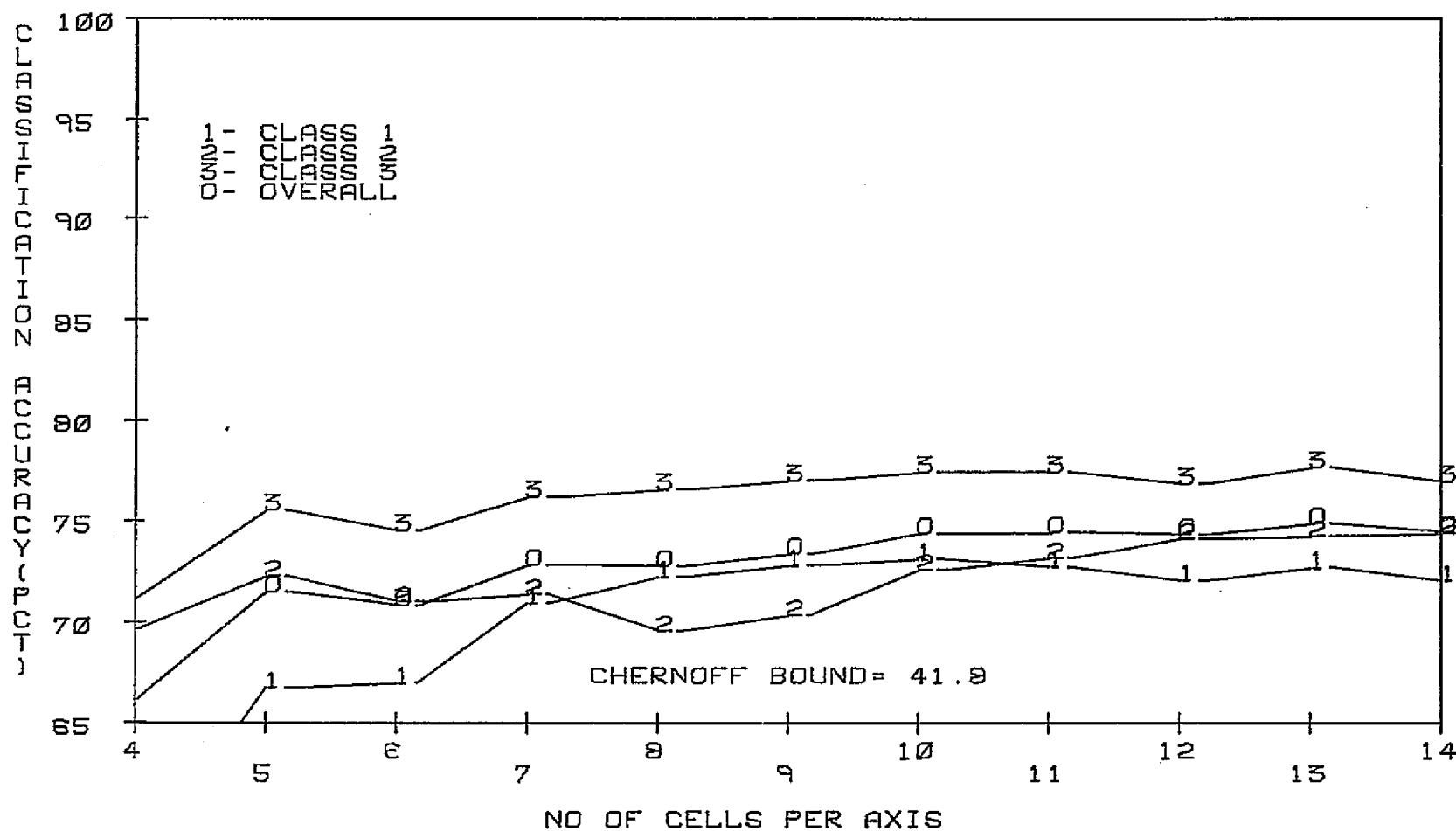


FIG. 4-38 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE MEAN. CASE 12

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 4-27 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP
ESTIMATION TECHNIQUE. CASE 13

G_s	$\hat{p}_{c \omega_1}$	$\hat{p}_{c \omega_2}$	$\hat{p}_{c \omega_3}$	$\hat{p}_c$
4	60.1	69.6	73.7	67.8
5	74.1	72.3	76.9	74.5
6	74.3	71.0	75.6	73.7
7	76.9	71.6	75.6	74.7
8	76.1	69.8	76.8	74.3
9	75.9	71.3	77.0	74.7
10	75.4	73.0	78.1	75.5
11	75.0	73.5	77.3	75.3
12	75.3	74.4	78.2	76.0
13	75.1	74.4	78.0	75.8
14	75.0	74.5	77.4	75.6

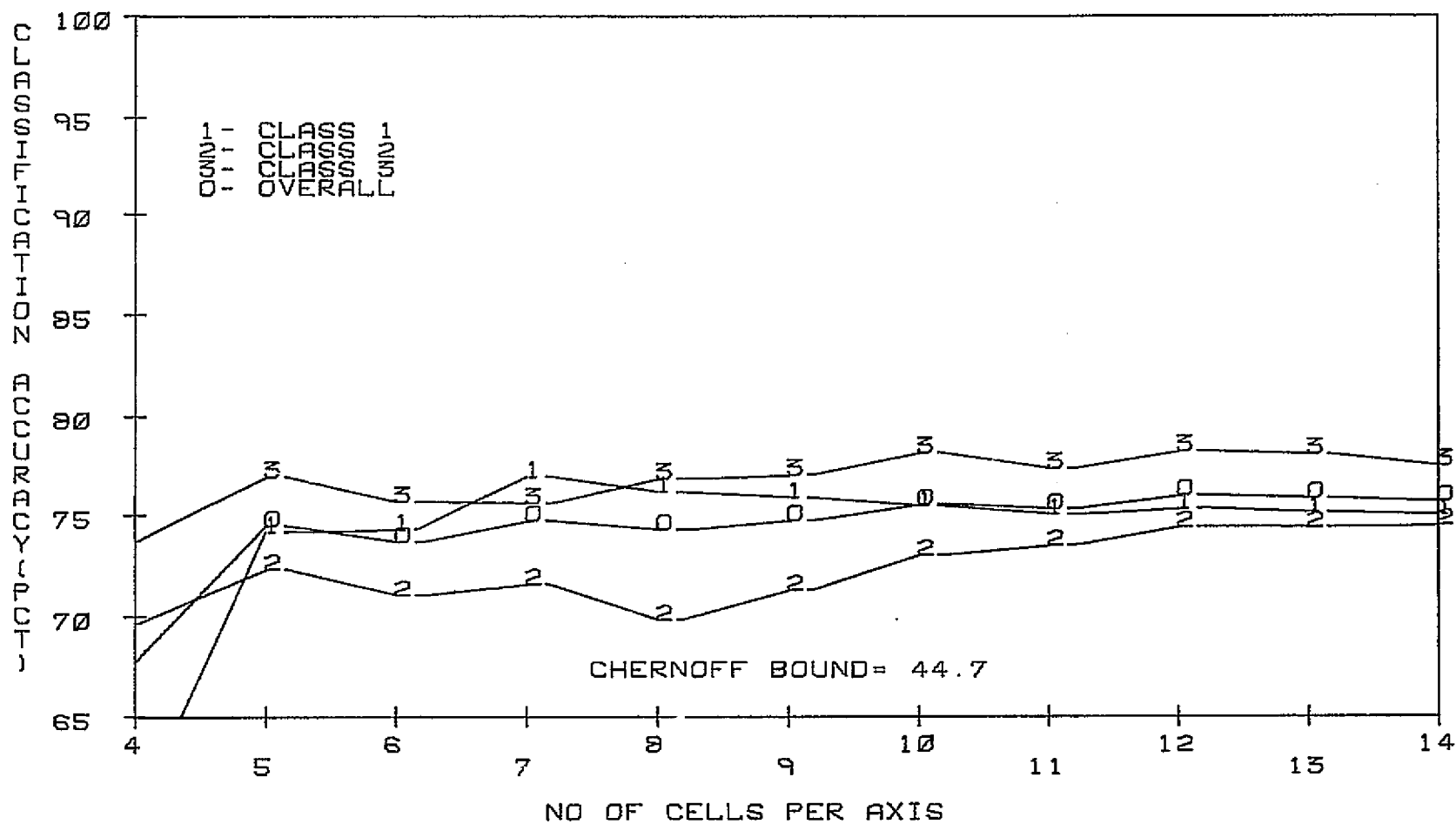


FIG. 4-39 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE MEAN. CASE 13

TABLE 4-28 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP
ESTIMATION TECHNIQUE. CASE 14

G_s	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{P}_c$
4	69.0	69.6	73.7	70.8
5	79.5	72.3	77.2	76.4
6	76.4	71.0	75.6	74.4
7	79.0	71.6	75.8	75.5
8	77.3	70.2	77.6	75.0
9	77.5	71.6	77.8	75.6
10	77.4	73.2	78.4	76.4
11	77.8	73.5	78.8	76.7
12	77.9	74.4	78.8	77.1
13	79.1	74.4	78.9	77.5
14	79.6	74.6	78.4	77.5

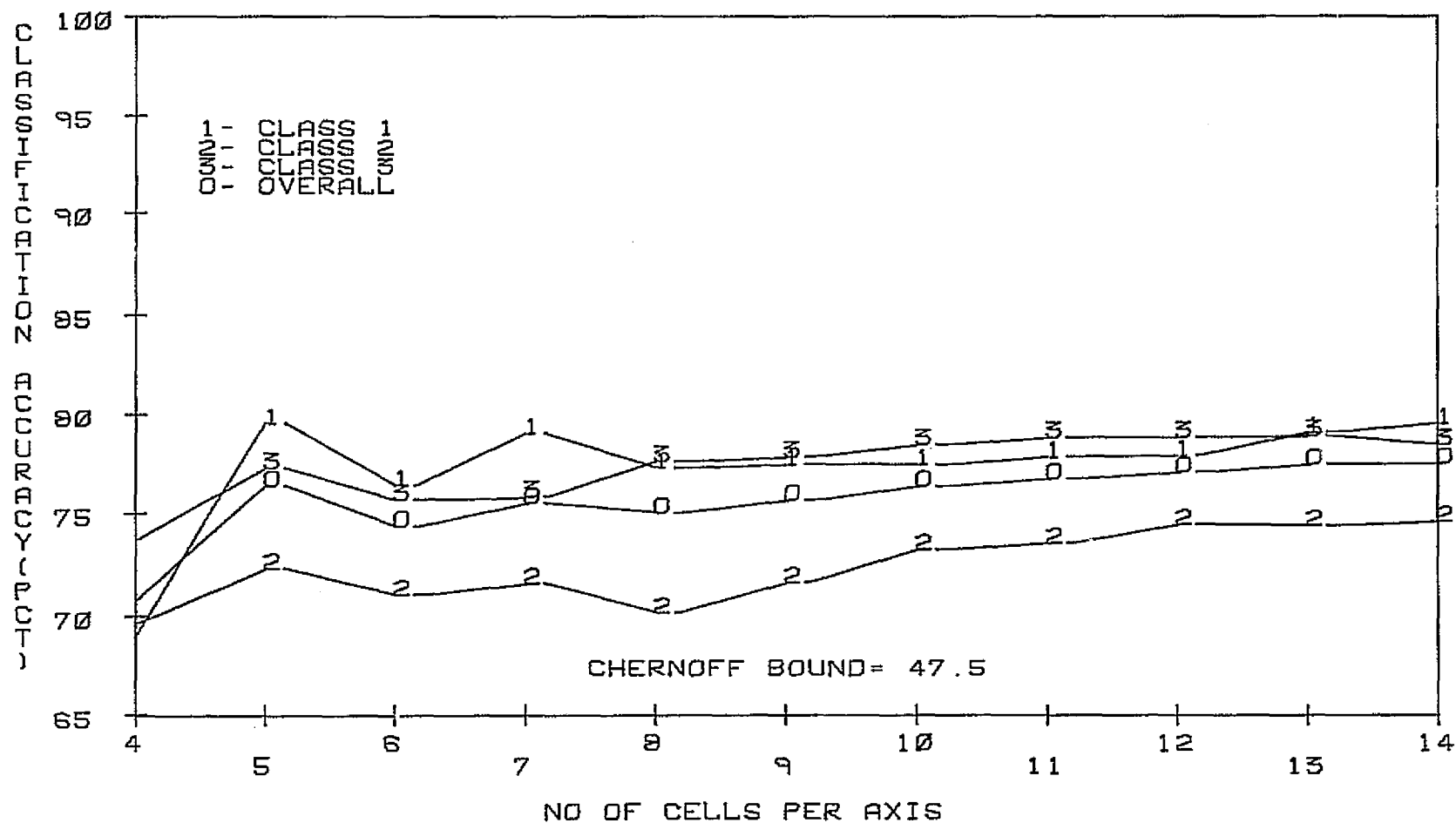


FIG. 4-40 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE MEAN. CASE 14

TABLE 4-29 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP
ESTIMATION TECHNIQUE. CASE 15

G_s	$\hat{P}_{C \omega_1}$	$\hat{P}_{C \omega_2}$	$\hat{P}_{C \omega_3}$	$\hat{P}_C$
4	74.7	69.6	75.1	73.1
5	80.9	72.3	77.2	76.8
6	77.5	71.2	76.4	75.0
7	80.0	72.1	75.9	76.0
8	78.5	71.0	78.3	75.9
9	80.3	71.8	78.5	76.9
10	80.8	73.3	79.4	77.8
11	81.7	73.6	79.6	78.3
12	82.8	74.5	79.7	79.0
13	82.6	74.5	79.3	78.8
14	82.3	74.7	79.2	78.7

ORIGINAL PAGE IS
OF POOR QUALITY

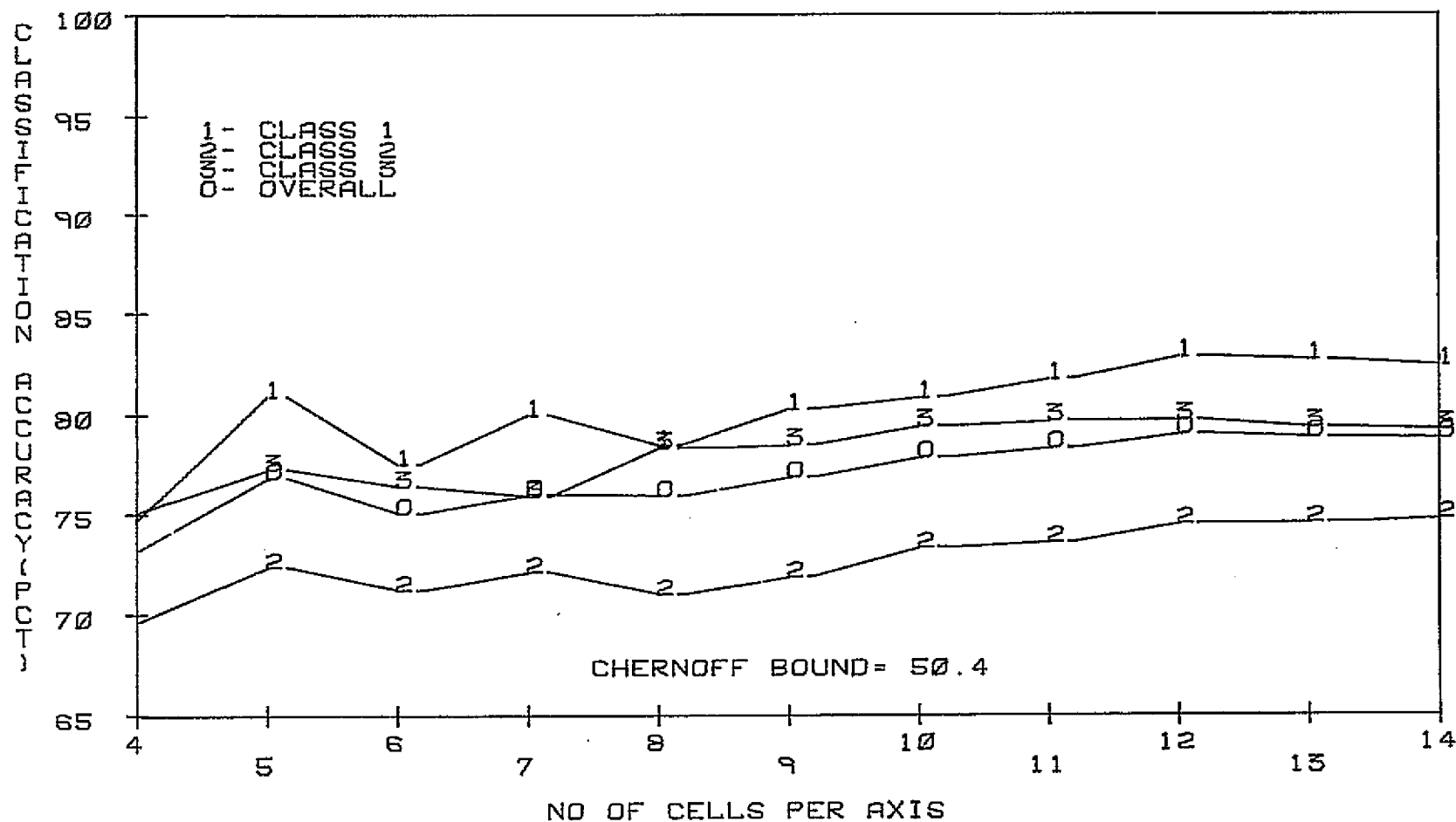


FIG. 4-41 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE MEAN. CASE 15

TABLE 4-30 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP
ESTIMATION TECHNIQUE. CASE 16

G_s	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{P}_c$
4	75.5	69.6	77.8	74.3
5	81.4	72.3	77.5	77.0
6	79.2	71.2	76.4	75.6
7	81.8	72.3	77.4	77.2
8	82.5	71.8	78.5	77.6
9	85.2	71.8	79.0	78.7
10	85.4	73.3	80.3	79.7
11	85.2	73.6	80.3	79.7
12	84.5	74.5	79.9	79.7
13	84.2	74.7	80.0	79.6
14	84.0	75.0	79.5	79.5

ORIGINAL PAGE IS
OF POOR QUALITY

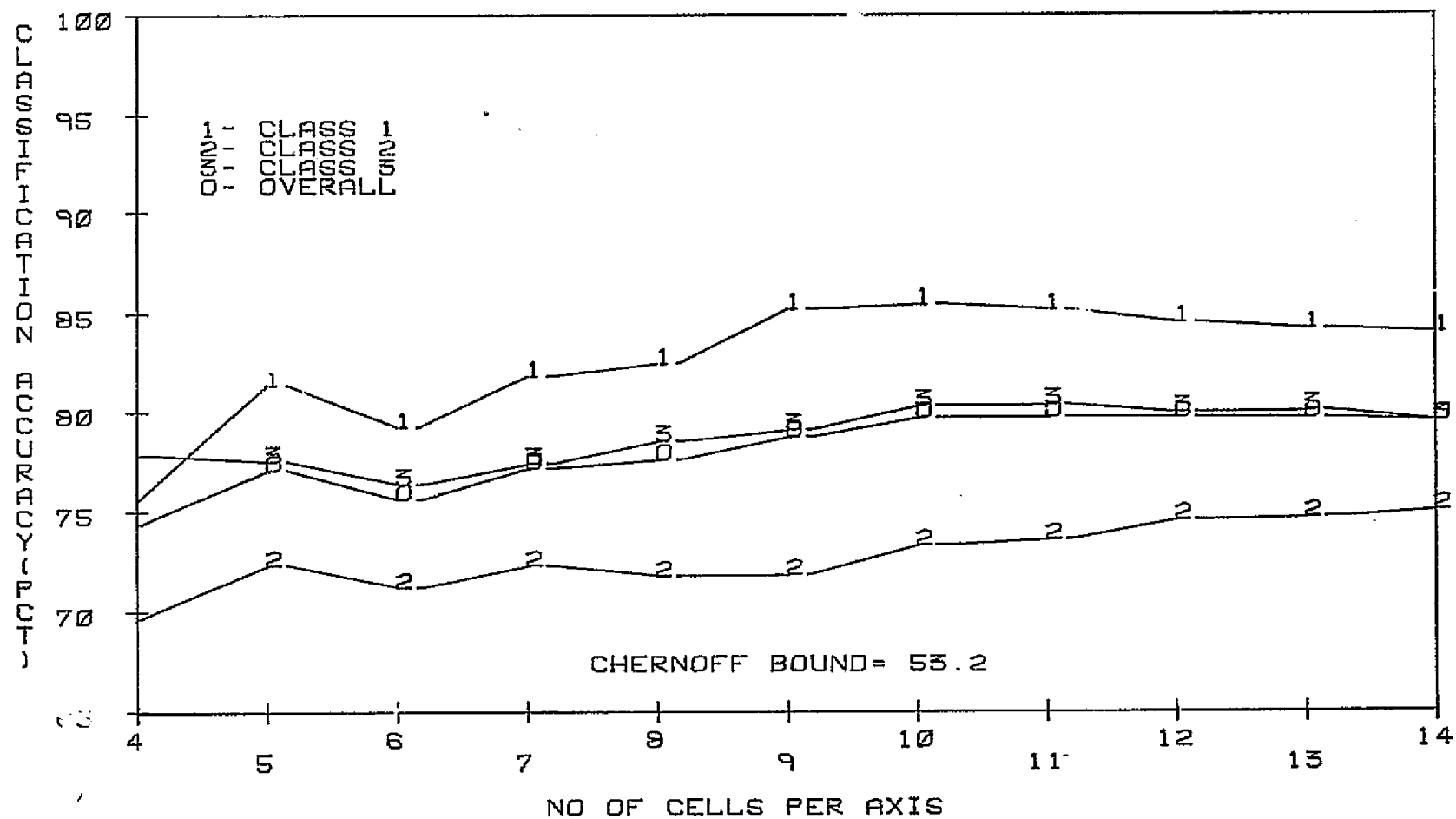


FIG. 4-42 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE MEAN. CASE 16

TABLE 4-31 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP
ESTIMATION TECHNIQUE. CASE 17

G_s	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{P}_c$
4	75.8	69.6	77.8	74.4
5	81.8	72.3	77.9	77.3
6	80.8	71.4	76.8	76.3
7	84.9	74.0	77.8	78.9
8	87.2	71.8	79.4	79.4
9	87.7	71.8	80.2	79.9
10	86.5	73.3	80.8	80.2
11	86.0	73.6	80.9	80.2
12	86.0	74.6	80.6	80.4
13	86.1	74.9	80.3	80.4
14	86.1	75.3	80.0	80.5

ORIGINAL PAGE IS
OF POOR QUALITY

0-3

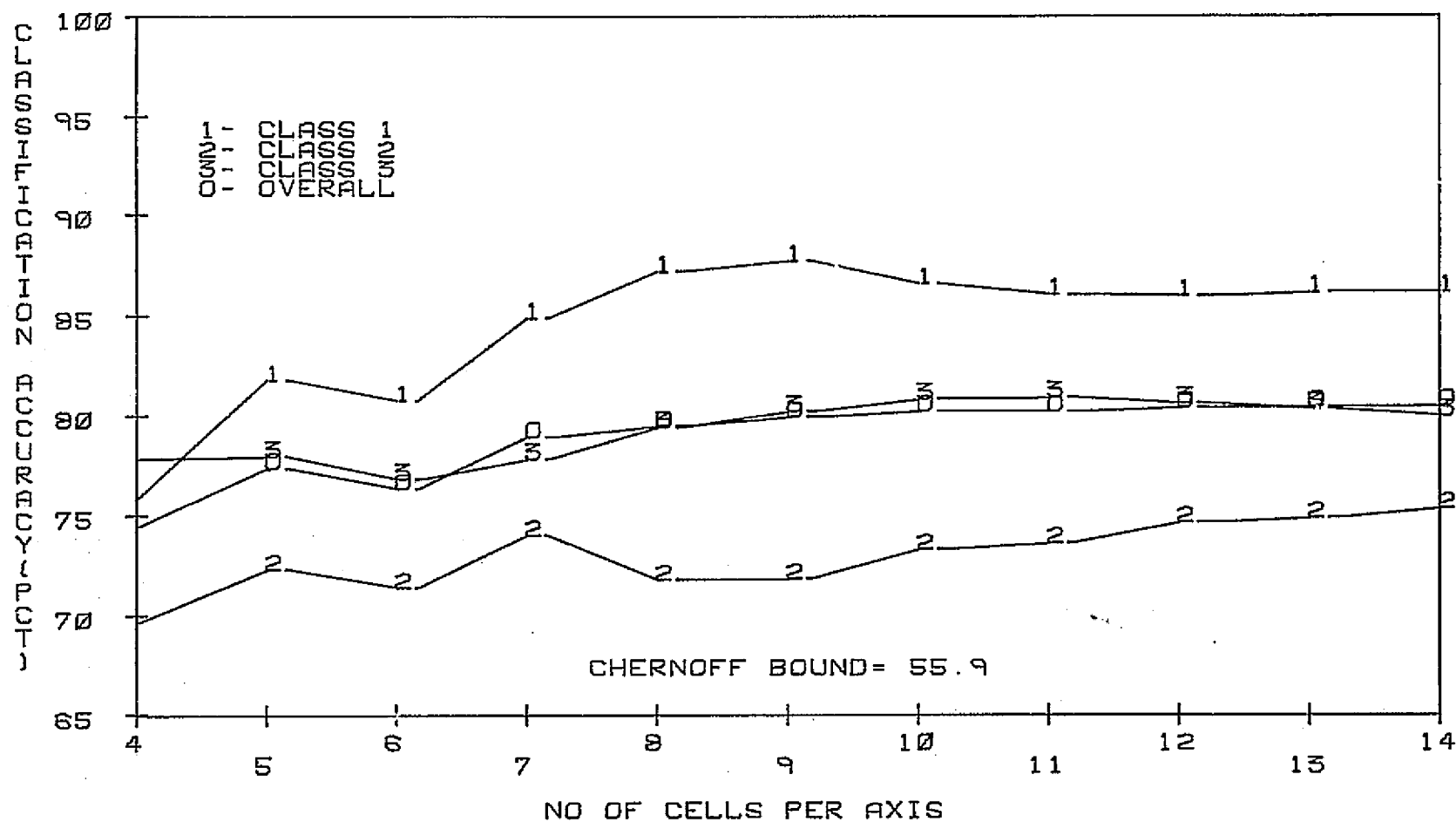


FIG. 4-43 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE MEAN. CASE 17

TABLE 4-32 PERCENT CLASSIFICATION ACCURACIES OBTAINED BY CSP
ESTIMATION TECHNIQUE. CASE 18

G_s	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{P}_c$
4	75.8	69.6 _m	77.8	74.4
5	83.2	72.4	78.5	78.0
6	83.6	71.7	77.7	77.7
7	89.6	74.0	78.6	80.7
8	88.6	71.8	79.4	79.9
9	88.2	71.9	80.2	80.1
10	87.0	73.3	81.9	80.8
11	87.8	73.7	81.3	81.0
12	88.0	74.9	81.0	81.3
13	89.8	75.3	80.8	82.0
14	89.9	75.9	81.0	82.3

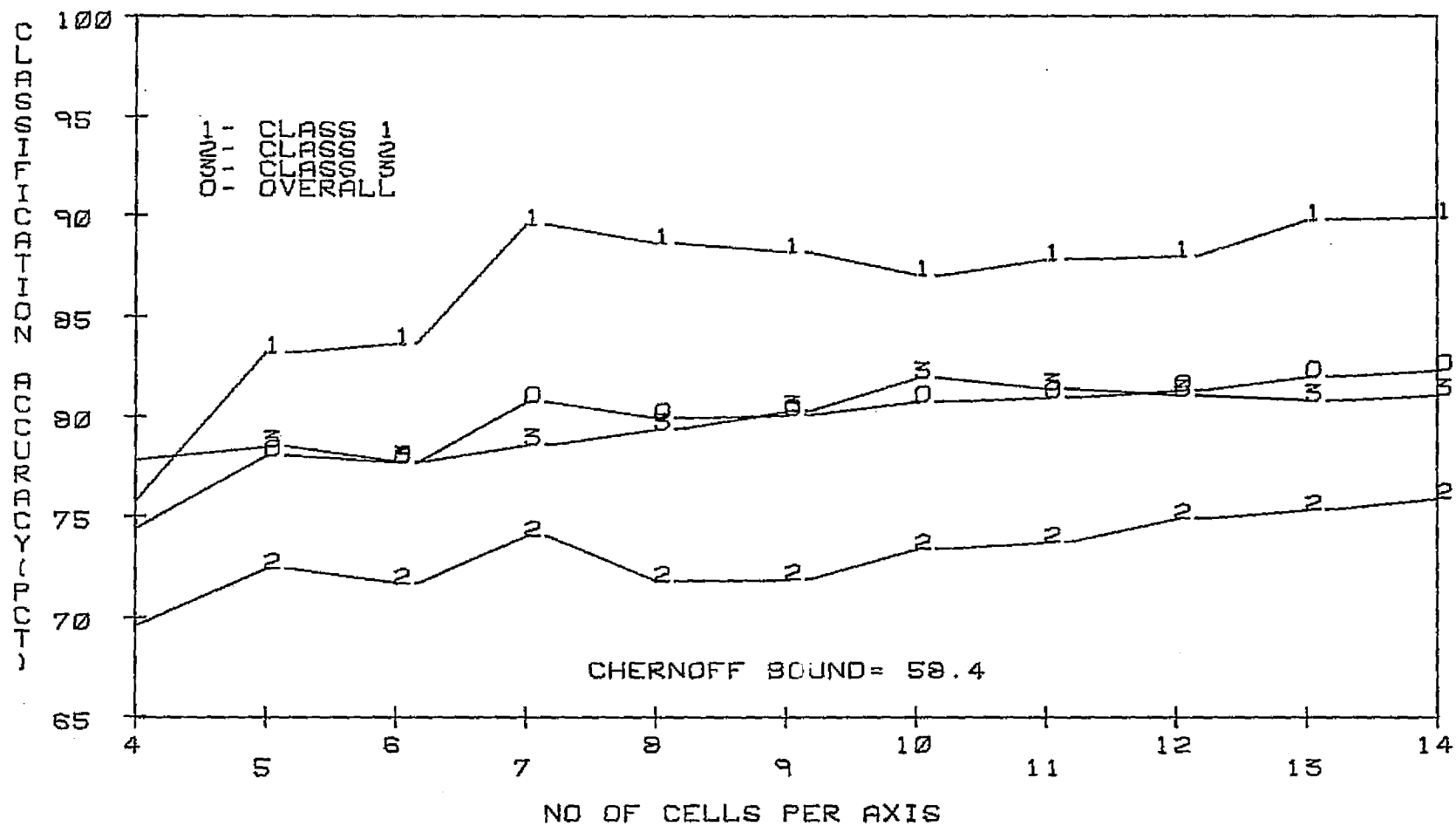


FIG. 4-44 CSP CLASSIFICATION ACCURACY ESTIMATE VS. GRID SIZE.
VARIABLE MEAN. CASE 18

of the results, however, the Monte Carlo error estimate is reported only for one sample size of about 6000. A large sample was chosen to assure a small bias and variance. Table 4-33 compares the CSP error estimates for the largest grid size with the corresponding MC estimation results. Cases 12 through 18 refer to the 7 cases listed in Table 4-25 with increasing variability. In examination of the results related to the variable mean case, all the CSP estimate properties are observed again; particularly evident is the generally rapid rise to a steady state value followed by a small magnitude oscillation. The curve corresponding to the Class 1 classification accuracy generally moves as expected. The separability increase by translation of μ_1 along x_1 also improves $\hat{P}_{c|\omega_2}$ and $\hat{P}_{c|\omega_3}$ but the improvement is not as great. $\hat{P}_{c|\omega_1}$ increased 18% from Case 12 to Case 18 while in the same range $\hat{P}_{c|\omega_2}$ and $\hat{P}_{c|\omega_3}$ improved 1.5% and 4.1%, respectively. The comparison of CSP and MC estimation results reveals that the differential between them is again small. For $\hat{P}_c$, the difference ranges from a high of 0.9% for Case 6 to a low of 0% for Case 4, Table 4-33.

4.2.3 Classification Error Estimation When the Bayes Rate is Known

Throughout this validation process the missing element has always been a fixed reference point in the form of a

TABLE 4-33 COMPARISON OF CSP AND MC PERCENT CLASSIFICATION ACCURACY.
VARIABLE MEAN.

Case	$\hat{P}_c \omega_1$		$\hat{P}_c \omega_2$		$\hat{P}_c \omega_3$		$\hat{P}_c$	
	CSP	MC	CSP	MC	CSP	MC	CSP	MC
1	72.1	72.5	74.4	74.4	76.9	76.9	74.5	74.6
2	75.0	76.2	74.5	74.6	77.4	78.3	75.6	76.3
3	79.6	78.8	74.6	76.4	78.4	78.7	77.5	78.0
4	82.3	81.3	74.7	74.8	79.2	80.1	78.7	78.7
5	84.0	84.9	75.0	75.7	79.5	80.1	79.5	80.7
6	86.1	87.1	75.3	74.9	78.0	80.1	79.8	80.7
7	89.9	89.0	75.9	76.0	81.0	80.9	82.3	82.0

known Bayes error. When only two classes are present in the data set, however, the desired quantity has been computed for up to an eight dimensional feature space [33]. The availability of this result provides two significant properties: (a) the reference error is sample-independent; (b) by working in a two dimensional subspace large grid sizes, impractical in higher dimensions, can be employed to observe the limiting behavior of the CSP algorithm and more importantly considerable insight to the geometry of the grid dynamics can be gained by actually displaying the domains of integration.

The variation of the probability of correct classification vs. grid size using the CSP estimation technique for a two dimensional feature space is shown in Fig. 4-45. The reported Bayes classification accuracies are superimposed on the plot and serve as asymptotes;

$$\begin{aligned} P_{c|w_1} &= 97.2\% \\ P_{c|w_2} &= 90.7\% \\ P_c &= 94.0\% \end{aligned} \quad (4-9)$$

The grid size ranges from a coarse 5 cells per axis to a very fine 75 cells per axis.

The behavior of the classification accuracy is somewhat different than the previous test cases. Two noticeable features are slower convergence and oscillations around

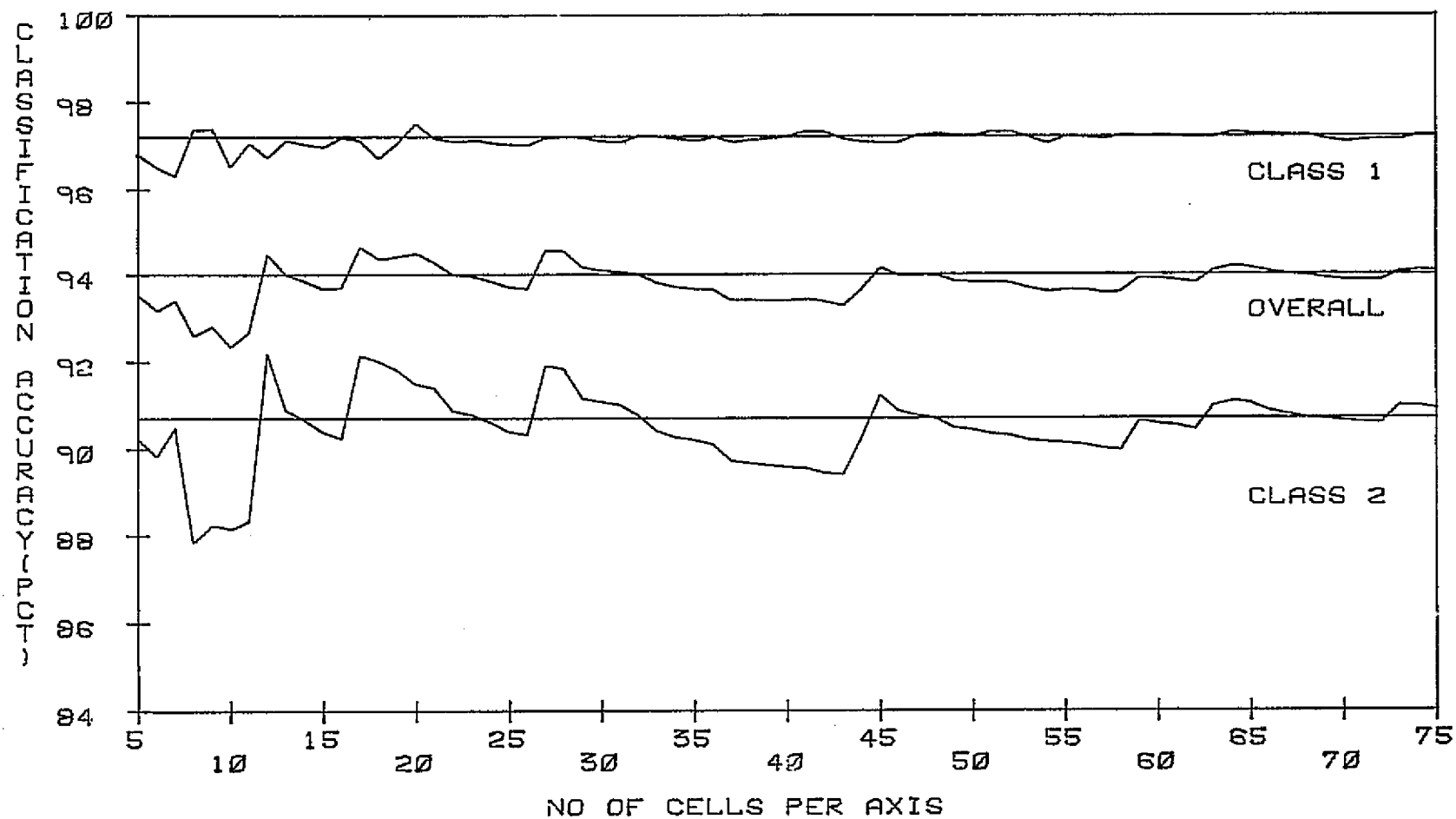


FIG. 4-45 COMPARISON OF BAYES AND CSP CLASSIFICATION ACCURACY ESTIMATES. TWO FEATURES

the asymptotes. The slower convergence can be traced to the property that the error estimate's variance is inversely proportional to the per axis cell number and exponentially to the dimensionality of the feature space. For example, on the basis of the total number of cells within the grid, 12 cells per axis in 2 dimensions where $\hat{P}_{c|\omega_1}$ has its highest surge, corresponds to somewhere between 3 and 4 cells per axis in 4 dimensional space and the corresponding numbers are only 8 and 9 for 75 two dimensional cells per axis. Another property not observed in the variable scatter or variable mean cases is the dissimilarity between the functional form of $\hat{P}_{c|\omega_1}$ and $\hat{P}_{c|\omega_2}$. In the 3 dimensional feature space examined before, all the estimates showed similar variational form with the grid size. In this case $\hat{P}_{c|\omega_2}$ exhibits periodic overshoots at 11-12, 15-16, 26-27, 43-44, etc. cells per axis. These oscillations are of the same nature as described in sec. 4.2.1. In this case, however, it is possible to get a close up of the actual estimation process. Consider the 15-16 jump. The two dimensional areas of integration are shown in Fig. 4-46. Take one scan line going through the ω_1 domain (dotted region). This line is marked with cell centers for three different grid sizes $G_s = 15, 16, 17$, Fig. 4-47. Denote two of these boundary cells by x_b and x'_b and let us follow their movement as grid size increases. For $G_s = 15$, $x_b = .54$, and $x'_b = 2.7$ are located inside and outside of the ω_2 domain. Recall that this domain is multiply connected. Therefore, the estimated

locations of the boundaries are in the intervals

$$0 < x \leq x_b = .54$$

(4-10)

$$2.14 \leq x \leq x'_b = 2.7$$

For $G_s = 16$, x_b and x'_b have moved to .52 and 2.58 respectively, and the boundary location is now narrowed to the domain

$$0 \leq x \leq x_b = .52$$

(4-11)

$$2.14 \leq x \leq x'_b = 2.58$$

x_b and x'_b are still one outside and one inside respectively.

The next grid size $G_s = 17$ is where $\hat{p}_{c|w_2}$ takes on a rapid jump. Now the boundary is determined to lie in the interval

$$.5 \leq x \leq 1$$

(4-12)

$$2.5 \leq x \leq 3.0$$

Comparing (4-12) with (4-11) establishes that the boundary must lie in the narrow interval

$$.5 \leq x \leq .52$$

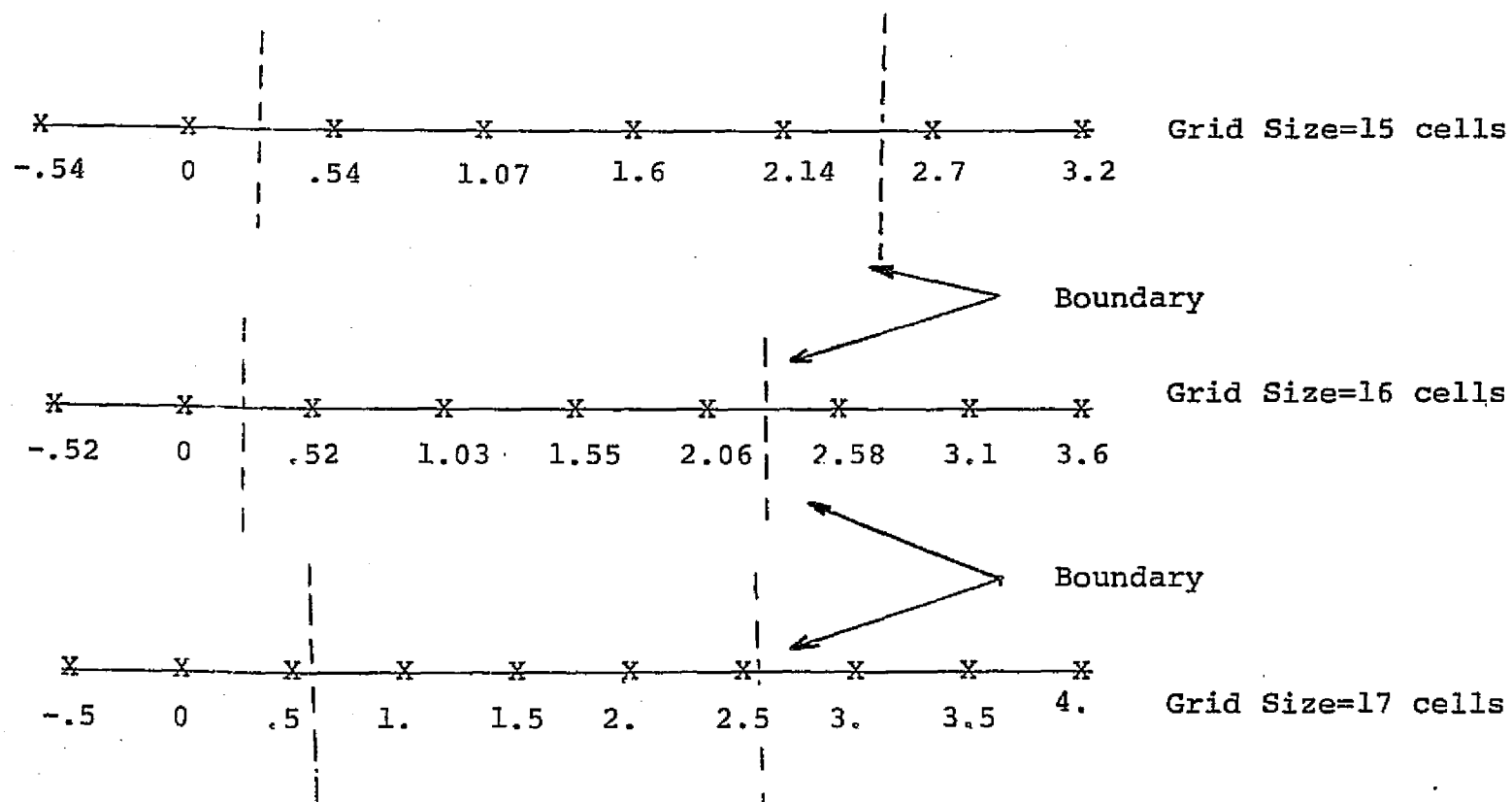
(4-13)

$$2.5 \leq x \leq 2.6$$

In this step, however, $x_b = .5$ has moved from outside to inside and $x'_b = 2.5$ made a similar move to outside of the integration domain. Recalling the discussion on the estimator's bias in sec. 2.4.1, shows that in this transition a net positive gain has occurred in the

```
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXX...XXX
XXXXXXXXXX....XXXX
XXXXXXXXXXXX...XXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
```

Fig. 4-46 Two Dimensional Partitioned Feature Space
... Class 1; xxx Class 2.



ORIGINAL PAGE IS
OF POOR QUALITY

Fig. 4-47 One Dimensional Grid Dynamics Illustration.

estimation of the volume under $f(\underline{X}|\omega_2)$, hence the surge in $\hat{P}_{C|\omega_2}$.

The comparison of the estimation results in 4 dimension with the one just obtained can shed some light on the effect of data dimensionality on the performance of the CSP algorithm. Using 4 features, the reported Bayes classification accuracies are [33]

$$\begin{aligned} P_{C|\omega_1} &= 97.4\% \\ P_{C|\omega_2} &= 95.0\% \\ P_C &= 96.2\% \end{aligned} \tag{4-14}$$

The results are shown in Fig. 4-48. Note that the functional form of this estimate is much more like the cases studied in a 3 dimensional feature space and the oscillation property is considerably less pronounced than its two dimensional counterpart due to a higher feature space dimensionality. The final values are all within 0.1% of the reported classification accuracy. In fact, considering that (4-14) is shown up to only 1 significant digit, the differential can be attributed to the round off factor and the estimates and their asymptote may well be identical.

4.3 Classification Accuracy Estimates Using Landsat and Aircraft MSS Data

The performance of the CSP estimation technique has been extensively investigated using simulated data. That

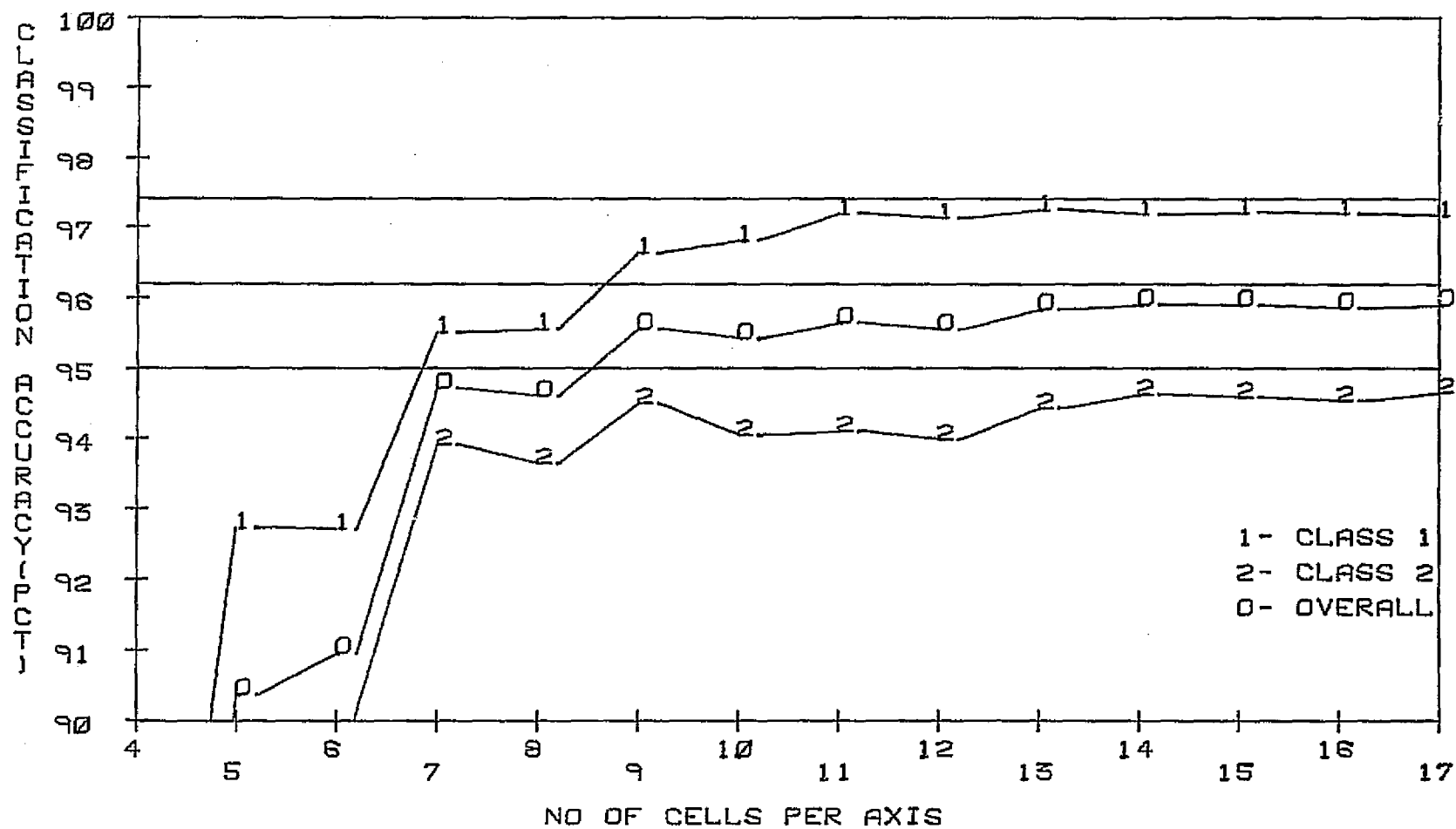


FIG. 4-48 COMPARISON OF BAYES AND CSP CLASSIFICATION ACCURACY ESTIMATES. FOUR FEATURES

study was intended to verify the proper operation of the algorithm. The actual application of this technique, however is the estimation of the probability of error in the classification of various cover types present in multispectral remotely sensed data. The currently operational Landsat-2 gathers information in 4 spectral bands. Therefore, the feature space is a moderate 4 dimensional domain where the CSP algorithm can effectively operate. In certain conditions some of the bands may be deemed redundant and thus a subset of the available 4 may be used. Three test regions were selected providing different numbers of cover types and classification error rates: Ogle county Illinois; Grant Couty Kansas; and Graham County Kansas.

The results of the parametric error estimator are compared with those of LARSYS, a data anlysis and classification technique developed at the Purdue University Laboratory for Applications of Remote Sensing. According to this algorithm a set of training fields is selected for each cover type based on ground truth information. These fields are then used to provide the necessary statistical input to an optimal Bayes classification such as (2-7). The entire frame of data is then classified by testing each pixel using (2-7). In order to obtain an estimate of the classification accuracy, a set of test fields is chosen and following the completion of the classification process, a count estimate such as (2-1) is computed for the misclassified

ORIGINAL PAGE IS
OF POOR QUALITY

Table 4-34 Percent Classification Accuracies
Obtained by CSP and LARSYS Algorithms,
Ogle County, IL

Class	No. of Samples	LARSYS%	CSP%
Corn	411	87.3	91.7
Soybean	224	90.6	91.3
Othere	217	94.0	90.6
Overall	852	90.7	91.2

Table 4-35 Percent Classification Accuracies
Obtained by CSP and LARSYS Algorithms,
Graham County, KAN

Class	No. of Samples	LARSYS%	CSP%
Baresoil	443	65.9	78.3
Corn	99	89.9	91.0
Pasture	1376	98.4	95.1
Wheat	459	94.8	93.9
Overall	2377	87.2	89.6

Table 4-36 Percent Classification Accuracies
Obtained by CSP and LARSYS Algorithms,
Grant County, KAN

Class	No. of Samples	LARSYS%	CSP%
AG1	793	52.3	59.3
AG2	445	75.8	73.3
AG3	134	90.3	88.8
Nonfarm	762	94.9	90.5
Wheat	930	82.7	79.7
Overall	3065	79.2	78.3

Table 4-37 Comparison of Percent Classification
Accuracies Obtained by CSP and LARSYS
Algorithms for Graham County Simulated data

Class	LARSYS%	CSP%	Difference
Baresoil	77.8	78.3	0.5
Corn	91.2	91.0	0.2
Pasture	95.3	95.1	0.2
Wheat	94.2	93.9	0.3
Overall	89.6	89.6	0.0

Table 4-38 LARSYS Classification Accuracies for
Three Realization of Graham County
Simulated Data.

Class	Random	Random	Random
	Start #1	Start #2	Start #3
Baresoil	77.0	79.7	79.0
Corn	91.2	92.1	91.0
Pasture	94.8	96.1	95.0
Wheat	94.0	94.2	94.8
Overall	89.2	90.5	90.0

samples. Frequently the training fields themselves are used in the performance calculations.

4.3.1 Ogle County, Illinois

This data is a portion of Landsat scene 1017-16093 acquired August 9, 1972, and has a LARS runtable entry of 72037806. Three training classes were used and classification was performed using 4 spectral lands; viz., channels 1 through 4. Table 4-34 shows both the classification accuracies obtained using the LARSYS point classifier and the CSP error estimation technique.

4.3.2 Graham County, Kansas

This data set is LACIE SRS segment 1018 and has a LARS runtable entry of 74078500. Channels 9 through 12 which are the acquisition corresponding to Landsat scene 1672-1644, were used. Four training classes were developed from 229 training fields. Results are tabulated in Table 4-35.

4.3.3 Grant County, Kansas

This data set is LACIE SRS segment 1036 and has a LARS runtable entry of 74027600. Channels 5 through 8 which are the acquisition corresponding to Landsat scene 1655-16512, were used in the classification study. Five training

CLASS...AG 1

TOTAL NUMBER OF SAMPLES... 793

HISTOGRAM(S)

CHANNEL 1 0.50 - 0.60 MICROMETERS

EACH * REPRESENTS 10 POINT(S).

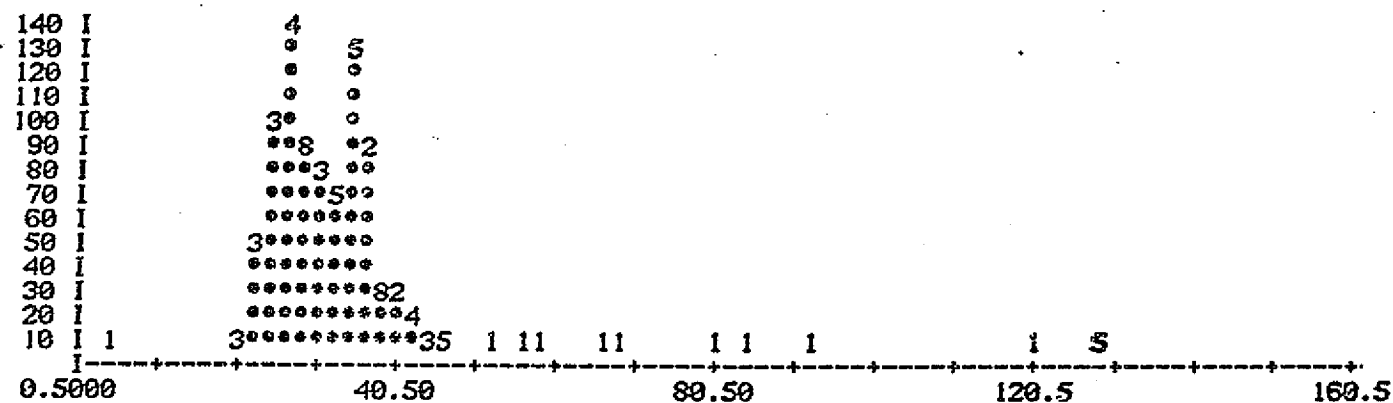


Fig. 4-49 Histogram of Actual Landsat Data, Grant Co. , KAN

ORIGINAL PAGE IS
OF POOR QUALITY

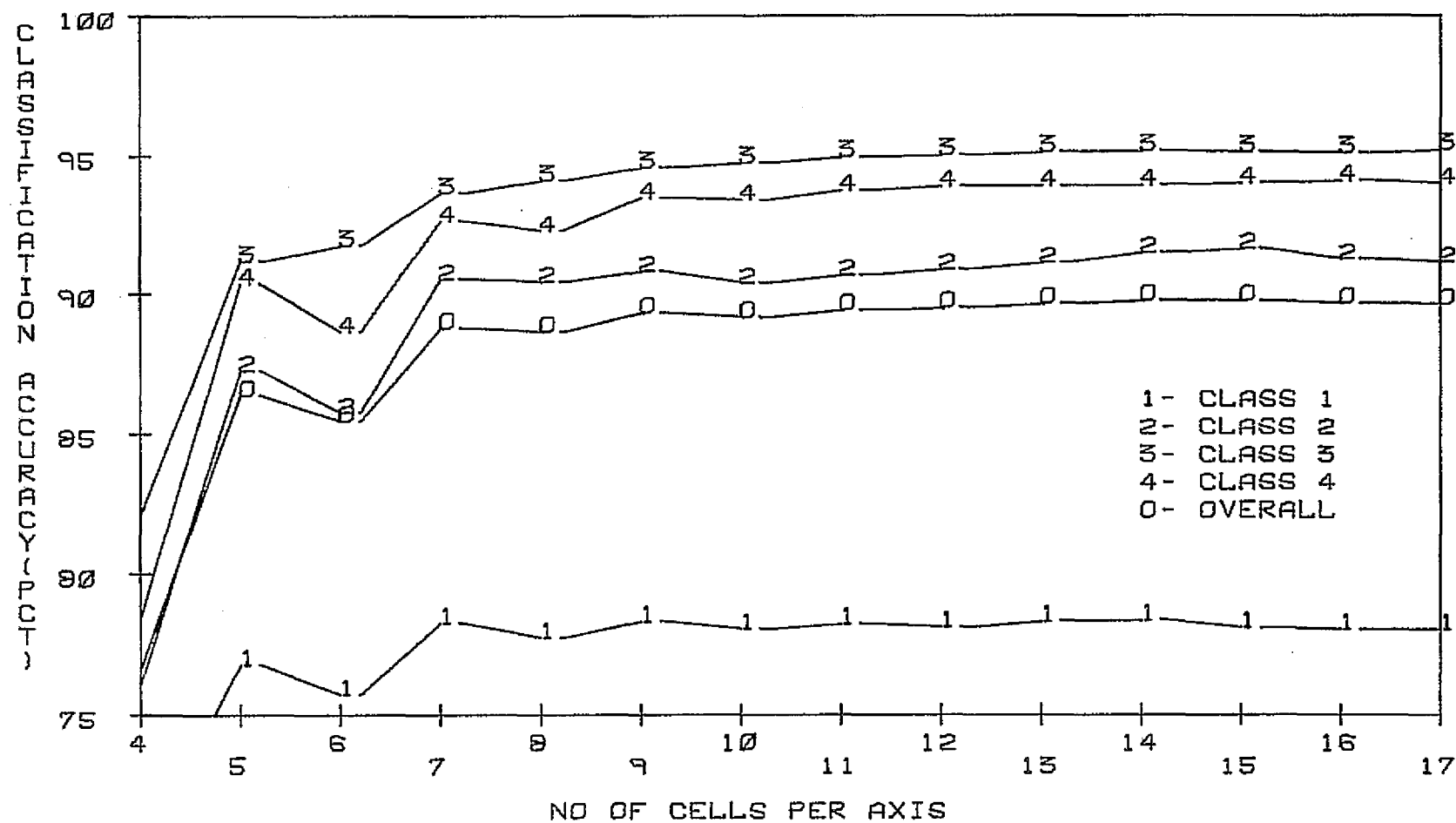


FIG. 4-50 CSP CLASSIFICATION ACCURACY ESTIMATE FOR GRAHAM CO. KAN.

classes were developed from 388 training fields. Results are tabulated in Table 4-36.

4.3.4 Discussion of the Results

Examining the results obtained here reveals that the performance of the CSP error estimator is consistent with that of sec. 4.2 using simulated data and closely matches MC classifier's output. As close as the CSP and LARSYS results may look, the differential in some cases is greater than the ones observed using artificial data. For example, in Graham County, bare soil is classified with 65.9% accuracy while the indicated theoretical value is 78.3, pasture is classified with 94.8% accuracy vs. the expected result of 95.1%. Similarly, in Grant County, AGI is classified with 52.5% accuracy vs. 59.3%. A possible explanation, initiated by examining the histogram of the actual data (Fig. 4-49), may lie in the validity of the assumption about the normality of the data under study or much more likely, the normality of the statistics of the training areas. In order to remove this element of uncertainty, artificial data was generated using the Graham County statistics. The simulated data was then reclassified using LARSYS. No attempt was made to keep the field sizes in the simulated data equal to those in the original set since the purpose of this step was to generate data having statistics as close to normal as possible. The new

classification accuracies are tabulated in Table 4-37.

The results are illuminating. While the bare soil was classified with a 65.9% accuracy in the actual data, in the simulated data this rate has risen to 77.8%, a gain of almost 12%, to put it within 0.5% of the result predicted by the CSP algorithm, similar observations can be made about other classes. The simulated data set being one realization of a stochastic process, makes the LARSYS results a random quantity. In order to make sure that Table 4-37 is not just one special case, the simulated data was re-generated three times using a different starting point for the pseudo-random number generator. The results shown in Table 4-38 confirm the preceding observations since the same close match exhibited before is repeated.

The results of the CSP error estimator are grid size dependent. Since variations in performance as a function of grid size were studied before, the classification accuracies reported here for the actual data are for a single G_s , usually around 12. For illustration purpose, the Graham County data was analyzed using a step wise grid employed before and the results are shown in Fig. 4-50. Note that the estimator exhibits the same properties observed repeatedly in the earlier studies.

4.4 Concluding Remarks

The performance of a multiclass multidimensional parametric Bayes classifier has been tested under widely

different conditions. In all cases, the results matched whatever reference point available. When that reference was the equivalent MC estimators, the CSP estimate was within 1% of it. Considering that an MC estimate is a random quantity, repeating the error estimation with another sample function of the process produces a different realization of the classification accuracy estimate. Therefore, an averaged MC estimator is well within the 1% maximum deviation from the CSP results. In Table 4-38 where three realizations of the MC estimator are computed, the overall classification accuracy estimates are 89.2%, 90.5%, and 90.0%. This compares with the fixed CSP estimate of 89.6%. Whereas the individual differences are 0.4%, 0.5%, and 0.4%, the averaged MC estimate, 89.9%, is 0.3% off. In fact, this difference may be reduced even further, if the number of MC estimates that are averaged is increased.

When the exact Bayes error rates were available, sec. 4.2.3, the CSP estimator provided essentially identical results. It has been shown that this algorithm has uniform performance with consistent systematic features throughout the test cases. One possible limitation emerged in that when hyperplanes parallel to the coordinate axes forming the boundaries of the feature space the CSP estimation technique performs poorly due to periodic high amplitude overshoots when there is a total shift of a considerable number of boundary cells from the outside to the inside

of integration domain and vice versa. However, this is expected to be an unlikely occurrence with real data.

In conclusion, it should be kept in mind that the classification error estimation algorithm developed here was not intended to provide a higher quality estimate than various random sampling techniques. The fact that it does so in many cases is only incidental but justifiable. The original goal was the development of an estimation algorithm dependent on the parameters of the problem alone. To that end the CSP estimation procedure has met the objectives.

CHAPTER 5

Experimental Evaluation of the MSS Spatial Model

This chapter is aimed at the validation and analysis of the scanner spatial model developed in chapter 3. Successful accomplishment of this task enables the integration of this model and the parametric Bayes error estimator as a complete set of tools for the evaluation of the performance of a MSS for any set of specified parameters. From Fig. 4-1 where the entire simulation process is depicted, it is seen that there are three phases involved. (i) Validation of the scanner characteristic function by comparing the output spectral covariance matrix of a convolution operator with that of the scanner linear system model. The input to the former is a simulated or real data set and to the latter is the statistical and spatial parameters of such a set. The results should closely match. (ii) Introduction of additive random Gaussian noise at the scanner input and output. (iii) Comparison of the probability of correct classification at the input to that of various output stages with noise power, scanner IFOV and data spatial structure as variables. Before embarking on the experiment, it is necessary to develop a suitable simulated data set.

5.1 Description of the Data Base

Stage II of the test data base simulation starts here. The checkout of the CSP error estimator required specification of the spectral characteristics of the data alone due to the fact that the spatial information is transparent to the Bayes spectral classification algorithm. The validation of the scanner model requires further conditioning of the stage I simulation output.

The 'white noise' property of the available test data although insignificant previously, would no longer be a realistic assumption about the multispectral data. In particular, the scanner's response is quite sensitive to the spatial structure of the input process (Fig. 35 through 3-12). It has been shown in sec. 3.3.2 that a Markov model closely approximates the spatial correlation of the multispectral data. Therefore, stage II of the simulation process consists of an additional transformation on the existing data base as a means of creating an exponential correlation property with any desired parameters. The technique to accomplish this task is formulated in the discrete domain in Appendix B. It is shown that filtering of a white noise process where the filter's PSF is a two dimensional one sided exponential,

$$h(x,y) = c_2 e^{-x/r_x} e^{-y/r_y} \quad x,y \geq 0 \quad (5-1)$$

generates a two dimensional random field with the adjacent sample and line correlation given by

$$\begin{aligned}\rho_x &= e^{-1/r_x} \\ \rho_y &= e^{-1/r_y}\end{aligned}\tag{5-2}$$

respectively. In addition to the correlation generating property, (5-1) inevitably alters the spectral structure of the input process. From (B-17) the output variance associated with any spectral band is given by

$$\sigma_g^2 = W^2(0,0) \left(\frac{\rho_x^{-2N_0} - 1}{\rho_x^{-2} - 1} \right) \left(\frac{\rho_y^{-2N_0} - 1}{\rho_y^{-2} - 1} \right) \sigma_f^2\tag{5-3}$$

where N_0 is the filter's PSF length in pixels, σ_f is the variance of the input process, σ_g the corresponding output quantity and $W(0,0)$ a quantity depending on r_x, r_y and N_0 . (5-3) approaches its continuous version for large N_0 which is

$$\sigma_g^2 / \sigma_f^2 = 1/4 r_x r_y\tag{5-4}$$

In chapter 3 it was pointed out that the magnitude of the variance reduction is large when a white noise process is transmitted through a MSS. Since the exponential filter is basically a linear system, the same property is observed in (5-4). For example, in order to generate a data set with the following somewhat typical correlation structure;

$$\begin{aligned}\rho_x &= 0.85 \\ \rho_y &= 0.75\end{aligned}\tag{5-5}$$

it requires that

$$\begin{aligned}r_x &= 6.15 \\ r_y &= 3.47\end{aligned}\tag{5-6}$$

From (5-4)

$$\sigma_g^2 / \sigma_f^2 = .012\tag{5-7}$$

and thus, the output variance is slightly over 1% of the input variance. This small fraction causes practical problems in generating the desired data set due to the finite dynamic range of the digital data on the storage medium. The representation of the problem in the discrete domain, however, provides the length of the filter as another variable to control the ratio expressed in (5-4). Let $N_0=5$ and r_x and r_y be as specified in (5-6). Then from (5-3)

$$\sigma_g^2 / \sigma_f^2 = 0.048\tag{5-8}$$

The resulting intersample and line correlations are now 0.65 and 0.53, respectively. The exponentially correlated data base is generated with adequate N/r_x and N/r_y ratio to closely approximate the continuously derived results.

5.2 Evaluation of the Scanner Characteristic Function

The scanner characteristic function, the transfer function that establishes a parametric/analytical relationship between the input and output of a multiband MSS, is the primary means by which various interactive processes within the scanner are studied. Like every other model developed so far, it is desirable to establish that near identical results are obtained using empirical techniques. In Fig. 4-1, this validation process is laid out. A white noise process with some prescribed statistics is generated and then conditioned to exhibit a specified pixel-to-pixel correlation. The actual data is transformed by a convolution operator having the PSF of the desired scanner and then the output statistics estimated. The statistics of the same input process are operated on by the scanner characteristic function and the output statistics directly computed. The comparison of the two resulting covariance and correlation matrices will produce the required result.

For this test the particular choice of the input statistics is relatively unimportant. Therefore, in order to use the data already available, Class 1 in Case 1 listed in Table 4-1 is selected as test data;

$$\underline{S}_{f_1} = \begin{bmatrix} 1.0 & 0.75 & 0.15 \\ & 1.0 & 0.45 \\ & & 1.0 \end{bmatrix} \quad (5-9)$$

The channel standard deviations are set at a large

$$\sigma_i = 30 \quad i = 1, 2, 3 \quad (5-10)$$

to cope with two successive variance reducing linear transformations. The variables of the problem are the scene correlation and the scanner IFOV. The IFOV is defined as the angle at the scanner subtended by a resolution element on the ground; e.g., 87 μ rad for the Landsat MSS. This definition when based on a Gaussian PSF is not unique. One convention that has been used [3] defines the IFOV as the angle between points where the PSF has dropped to half its peak amplitude, Fig. 5-1. Throughout this chapter the definition adopted is such that IFOV and characteristic length, r_0 , are identical.

Scanner systems with different resolution capabilities and different signal sampling intervals produce images with different adjacent pixel separation. In order to eliminate the dependency of the problem formation on the actual physical distance between each resolution element, the spatial parameters are normalized to that quantity and thus many of the results are on a per pixel basis. Later in the experiment alternate conventions are defined based on the particular problem under study. According to this definition, pixel separation in effect is unity. This assumption is particularly relevant in the simulation

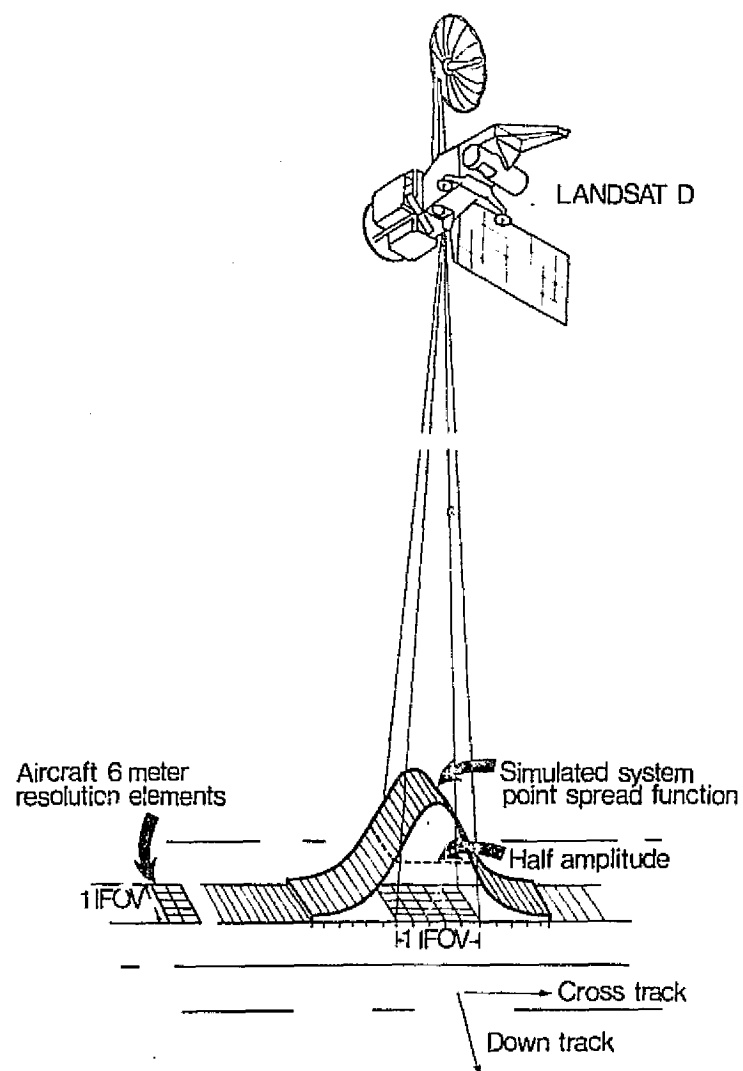


Fig. 5-1 Conceptual Illustration of a Picture Element Viewed from the Satellite.

ORIGINAL PAGE IS
OF POOR QUALITY

stages of the experiment since data is artificially generated and one can assign any desired quantity to the samples and lines separation interval.

In order to experimentally verify the theoretical variations of the scanner characteristic function a test data set is required. The adjacent sample correlation is scanned from 0.6 to 0.9 with an increment of 0.05 while the adjacent line correlation is kept at 0.7 for the first set and 0.8 for the second. For test purposes two sets of scanner PSF's with $r_o = 1$ and 4 pixels are selected. The particular choice of these parameters are again somewhat arbitrary. An attempt was made, however, to make the selections realistic in terms of practical systems. For each adjacent sample correlation, adjacent line correlation and the scanner IFOV, the ratio of the output variance to the input variance is experimentally determined and the results superimposed on the theoretical plot of the characteristic function vs. scene correlation. This is done for one spectral band and the results are shown in Fig. 5-2 and 5-3. The percent difference between the theoretical and experimental characteristic functions, W_i and $\hat{W}_i$ is expressed in Tables 5-1 through 5-4, where W_i is the functional value for the i th sample-to-sample correlation in the corresponding table.

Examination of the error terms and the accompanying figures indicates a relatively close match between the two

TABLE 5- 1 EXPERIMENTAL AND THEORETICAL SCANNER CHARACTERISTIC
FUNCTION. IFOV= 1, ADJACENT LINE CORRELATION=0.70

ρ_x	$\hat{W}_i$	W_i	$e_i\%$
0.60	0.56	0.54	3.70
0.65	0.58	0.56	3.60
0.70	0.60	0.59	1.70
0.75	0.62	0.62	0.0
0.80	0.63	0.65	3.10
0.85	0.65	0.68	4.40
0.90	0.65	0.70	7.10

TABLE 5- 2 EXPERIMENTAL AND THEORETICAL SCANNER CHARACTERISTIC
FUNCTION. IFOV= 4, ADJACENT LINE CORRELATION=0.70

ρ_x	$\hat{W}_i$	W_i	$e_i\%$
0.60	0.12	0.14	14.30
0.65	0.13	0.16	18.70
0.70	0.15	0.18	16.60
0.75	0.18	0.20	10.00
0.80	0.20	0.23	13.00
0.85	0.23	0.27	14.80
0.90	0.28	0.31	9.70

TABLE 5- 3 EXPERIMENTAL AND THEORETICAL SCANNER CHARACTERISTIC
FUNCTION. IFOV= 1, ADJACENT LINE CORRELATION=0.80

ρ_x	$\hat{W}_i$	W_i	$e_i \%$
0.60	0.61	0.56	8.90
0.65	0.62	0.62	0.0
0.70	0.65	0.65	0.0
0.75	0.66	0.68	2.90
0.80	0.68	0.71	4.20
0.85	0.68	0.74	8.10
0.90	0.70	0.78	10.20

TABLE 5- 4 EXPERIMENTAL AND THEORETICAL SCANNER CHARACTERISTIC
FUNCTION. IFOV= 4, ADJACENT LINE CORRELATION=0.80

ρ_x	$\hat{W}_i$	W_i	$e_i \%$
0.60	0.15	0.18	16.60
0.65	0.17	0.21	19.00
0.70	0.20	0.24	16.60
0.75	0.23	0.27	14.80
0.80	0.26	0.31	16.10
0.85	0.31	0.35	11.40
0.90	0.37	0.41	9.70

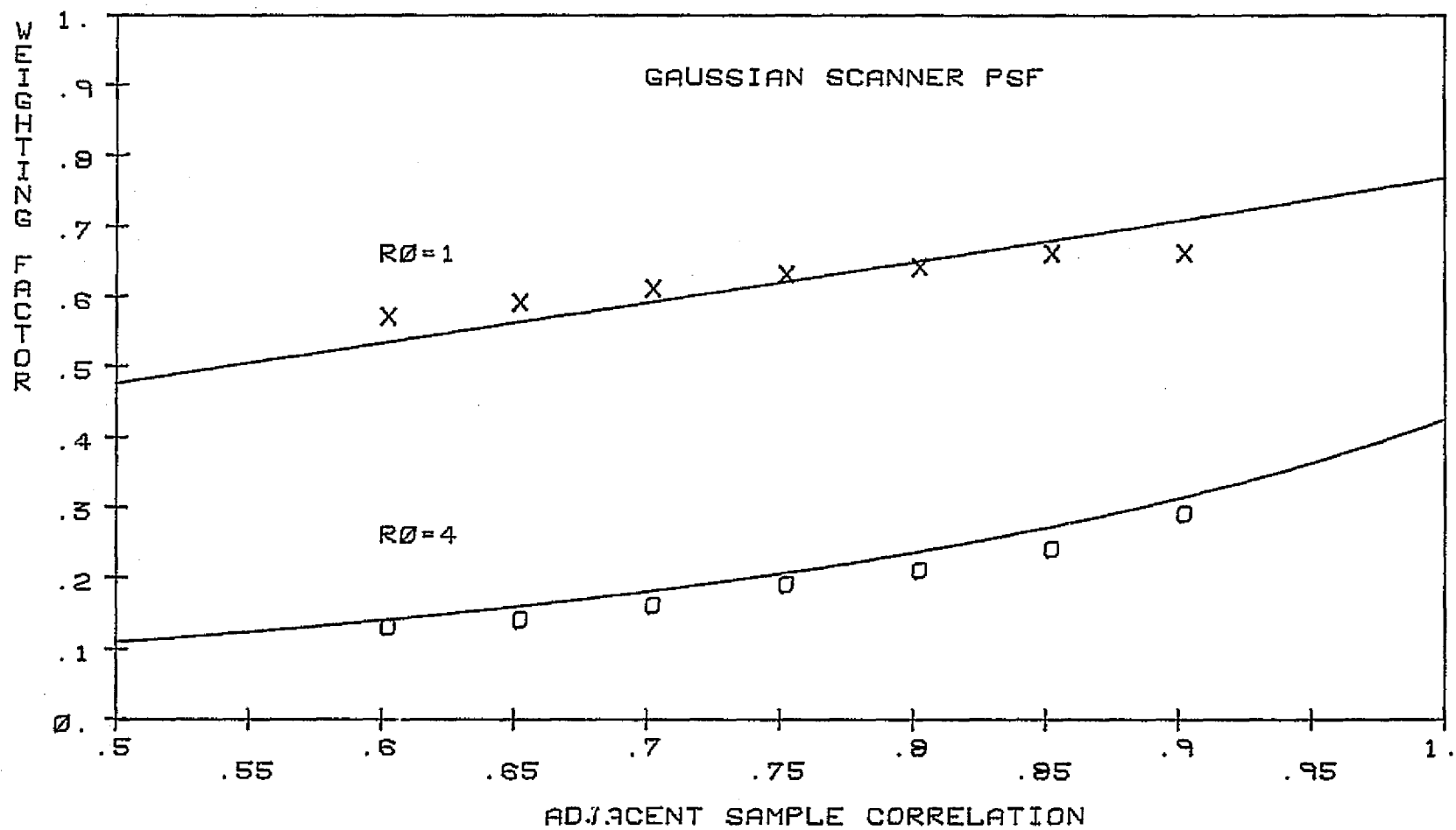


FIG. 5-2 EXPERIMENTAL AND THEORETICAL MSS CHARACTERISTIC FUNCTION
ADJACENT LINE CORRELATION = .7

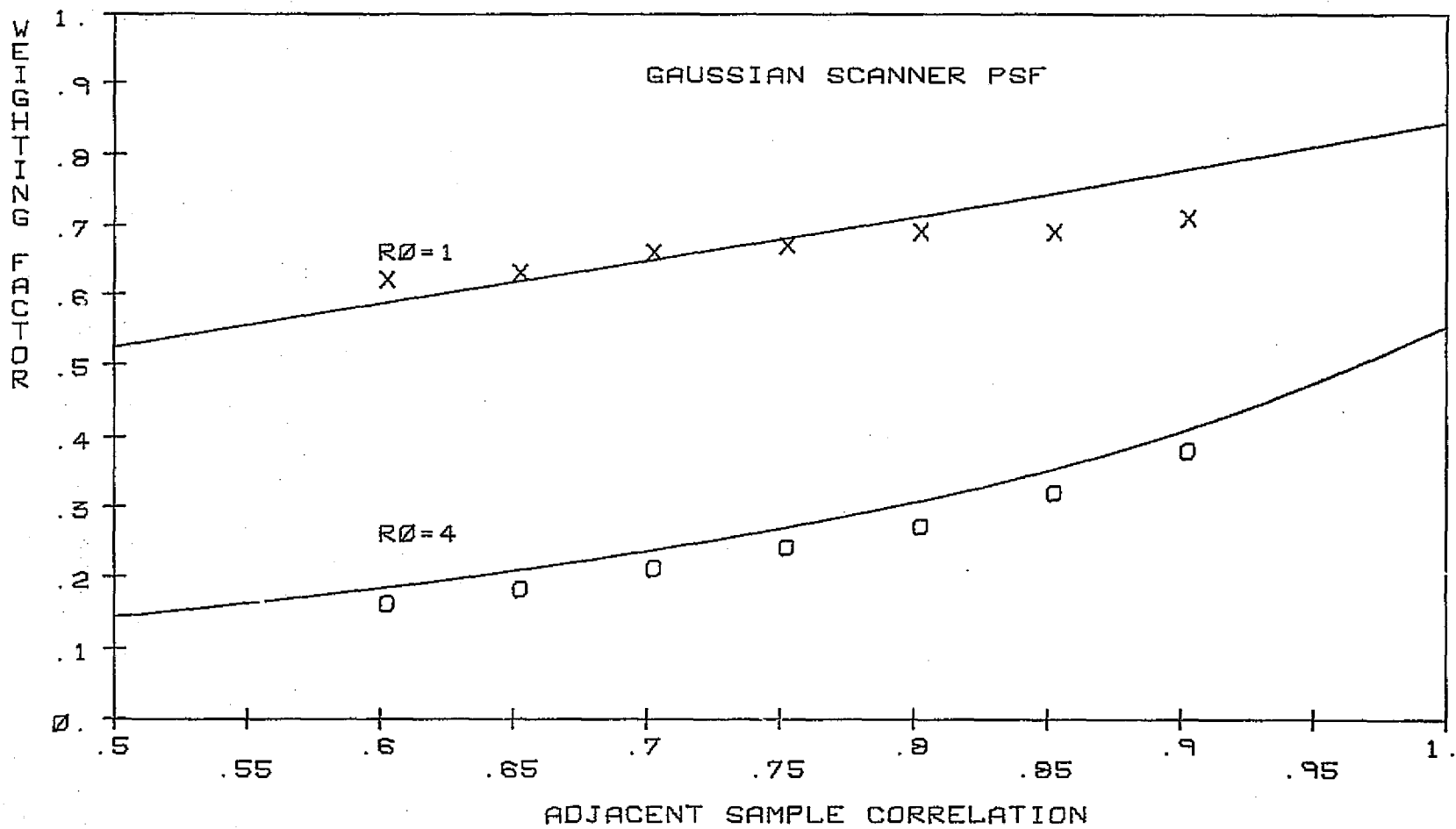


FIG. 5-3 EXPERIMENTAL AND THEORETICAL MSS CHARACTERISTIC FUNCTION
ADJACENT LINE CORRELATION- .8

ORIGINAL PAGE IS
OF POOR QUALITY

independently derived functions. When $\rho_y = 0.7$ and $r_o = 1$ pixel, the percent error ranged from a high of 7.1% at $\rho_x = 0.9$ to 0% at $\rho_x = 0.75$ for an average of 3.4%. For $r_o = 4$ pixels the percent error ranged from a high of 18.7% at $\rho_x = 0.65$ to 9.7% at $\rho_x = 0.9$ for an average of 13.9%. The explanation for a higher discrepancy between the theoretical and experimental values of the latter case can be attributed to the inherent error of a discrete testing of an essentially continuous phenomenon. Thile this error is always present, under certain unfavorable conditions may become significant. In this case a large IFOV dictates the choice of PSF with a considerably greater number of samples in order to satisfactorily approximate its continuous counterpart. This in turn requires a larger size data base and accompanying increase in computation time. The last factor was the main constraint that limited the PSF's length and contributed to the increase in deviation from the theoretical result. This factor notwithstanding, Fig. 5-2 and 5-3 show a very acceptable harmony between the two results and provide substantial evidence for the validity of the analytic scanner characteristic function.

This validation was accomplished in the context of variances alone. That this is not a special case is easily concluded from the property that the output cross-channel spectral correlation coefficients are simply ratios

of one or several appropriate characteristic functions. Having tested its building block, the experimental verification of the entire spectral correlation matrix is implicitly accomplished.

By means of the above evaluation the parametric models developed for the analysis of the MSS performance is accomplished. Therefore, unless stated otherwise, all the results obtained hereafter will be based entirely on the statistical properties of the populations, scanner parameters etc. and no data bases, simulated or measured, will be employed.

5.3 MSS and Classifiability of the Multispectral Data

A major application of the various parametric models and methods developed during this study is in determining the interactions among the MSS system parameters on a data-independent basis. Having experimentally verified the validity of the models, such evaluation of the performance of a multispectral scanner is feasible. In any system analysis the definition of an index of performance is a basic requirement. When the system is a MSS in a remote sensing data gathering package, the accuracy by which various populations present in the final data set are classified primarily determines the degree of success of the initial design. Therefore, throughout this chapter the objective is to observe the probability of current classification at various stages of the MSS, Fig. 5-4,

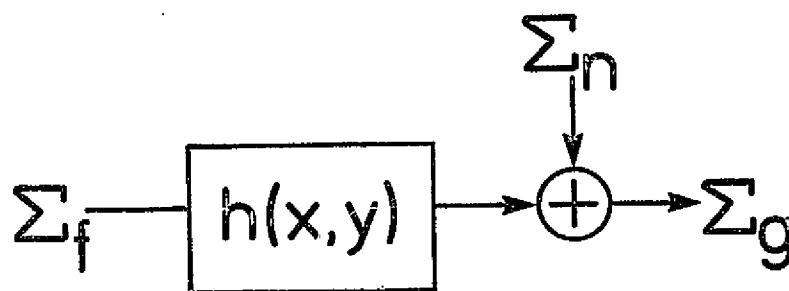


Fig. 5-4 A Statistical Illustration of the MSS Model.

ORIGINAL PAGE IS
OF POOR QUALITY

and monitor its variations with the SNR, scanner IFOV and spatial correlation of the scene. A Gaussian PSF is employed unless stated otherwise.

5.3.1 Classification Accuracies at the MSS Output: No Noise

The test statistics is Case 1 in Table 4-1 containing 3 classes with 3 features. The input to the scanner is a spatially correlated data set with an adjacent sample correlation ranging from 0.5 to 0.95 in steps of 0.05. For each ρ_x a corresponding ρ_y is computed on the following basis. The sampling of the analog Landsat data is such that the ratio of the ground distance between the cross-track pixels to that of along-track is about 0.7. Since the adjacent pixel correlations along these two directions are equal in a continuous model, it follows that if

$$\rho_x^\tau = e^{-a\tau} \quad (5-11)$$

$$\rho_y^\eta = e^{-b\eta} \quad \tau, \eta = 0, 1, 2, \dots \quad (5-12)$$

then

$$a = 0.7b \quad (5-13)$$

therefore

$$\rho_y^\tau = \rho_x^{\tau(10/7)} \quad (5-14)$$

With the input statistics defined as above, 10 cases are obtained and for each case r_0 is varied from 1 to 8 pixels

and an output classification accuracy is estimated for each combination of the scene correlation and scanner IFOV. Fig. 5-5 through 5-14 and Tables 5-5 through 5-14 show the variation of the output probabilities of correct classification as a function of IFOV. 13 cells per axis are used in the CSP error estimation algorithm.

The variations of the output probabilities of correct classification are in complete agreement with those projected by the characteristic function. The most notable feature is the inverse relationship between the scene spatial correlation and the slope of $\hat{P}_c|_{w_i}$ vs. IFOV at the output. When the scene is spatially highly uncorrelated such as Fig. 5-5, $\hat{P}_c$ gained 16.2% by increasing the IFOV from 1 to 2 pixels wide, whereas, the same increase in IFOV produced a gain of 9.7% for $\rho_x = 0.6$, 6.7% for $\rho_x = 0.7$, 3.3% for $\rho_x = 0.8$ and only 0.9% when $\rho_x = 0.95$. This behavior can be predicted from the variations of W_s vs ρ_x . Referring to Fig. 3-5 through 3-12 where W_s is plotted, it is observed that the one step reduction in input variance gets progressively smaller toward higher scene correlations. For the test case under study where any reduction of the class variances along a feature axis can contribute to increased separability, the aforementioned property of W_s accounts for the changing slope of $\hat{P}_c|_{w_i}$ over the ensemble of the scene spatial correlations.

TABLE 5- 5 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 ADJACENT SAMPLE CORRELATION=0.50

IFOV	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{P}_c$
1	68.9	74.2	76.6	73.2
2	82.4	86.0	84.9	84.4
3	91.5	94.3	92.7	92.8
4	96.5	97.8	97.0	97.1
5	98.7	99.2	98.9	99.0
6	99.6	99.8	99.7	99.7
7	99.9	99.9	99.9	99.9
8	99.9	99.9	99.9	99.9

ORIGINAL PAGE IS
 OF POOR QUALITY

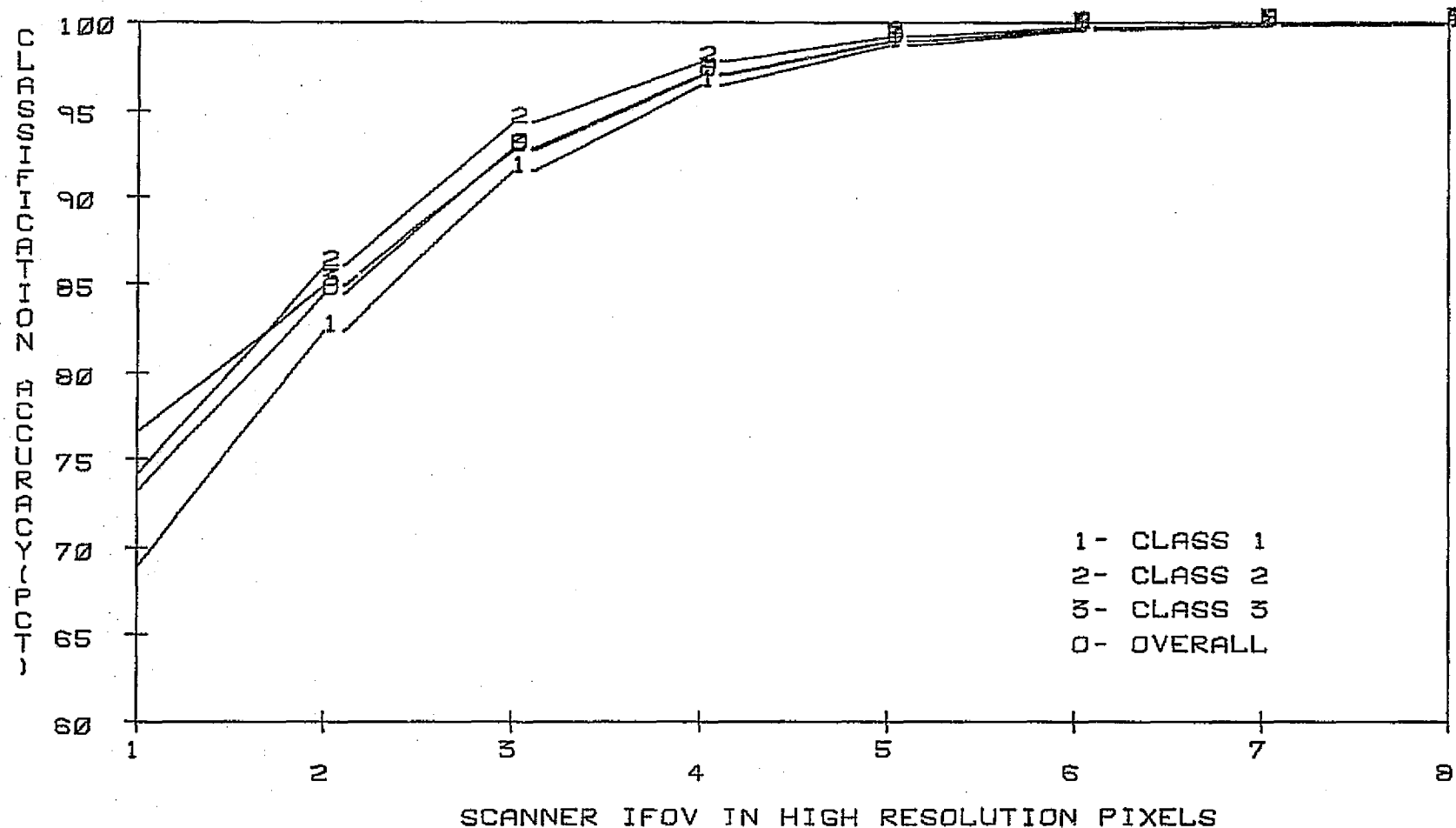


FIG. 5-5 SCANNER OUTPUT CLASSIFICATION ACCURACY VS. IFOV
ADJACENT SAMPLE CORRELATION = .5

TABLE 5- 6 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 ADJACENT SAMPLE CORRELATION=0.55

IFOV	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{P}_c$
1	66.5	70.6	75.6	70.9
2	78.3	83.7	82.6	81.6
3	87.7	91.6	89.8	89.7
4	94.2	95.9	94.8	95.0
5	97.3	98.4	97.7	97.8
6	98.9	99.4	99.1	99.1
7	99.6	99.8	99.7	99.7
8	99.8	99.9	99.9	99.9

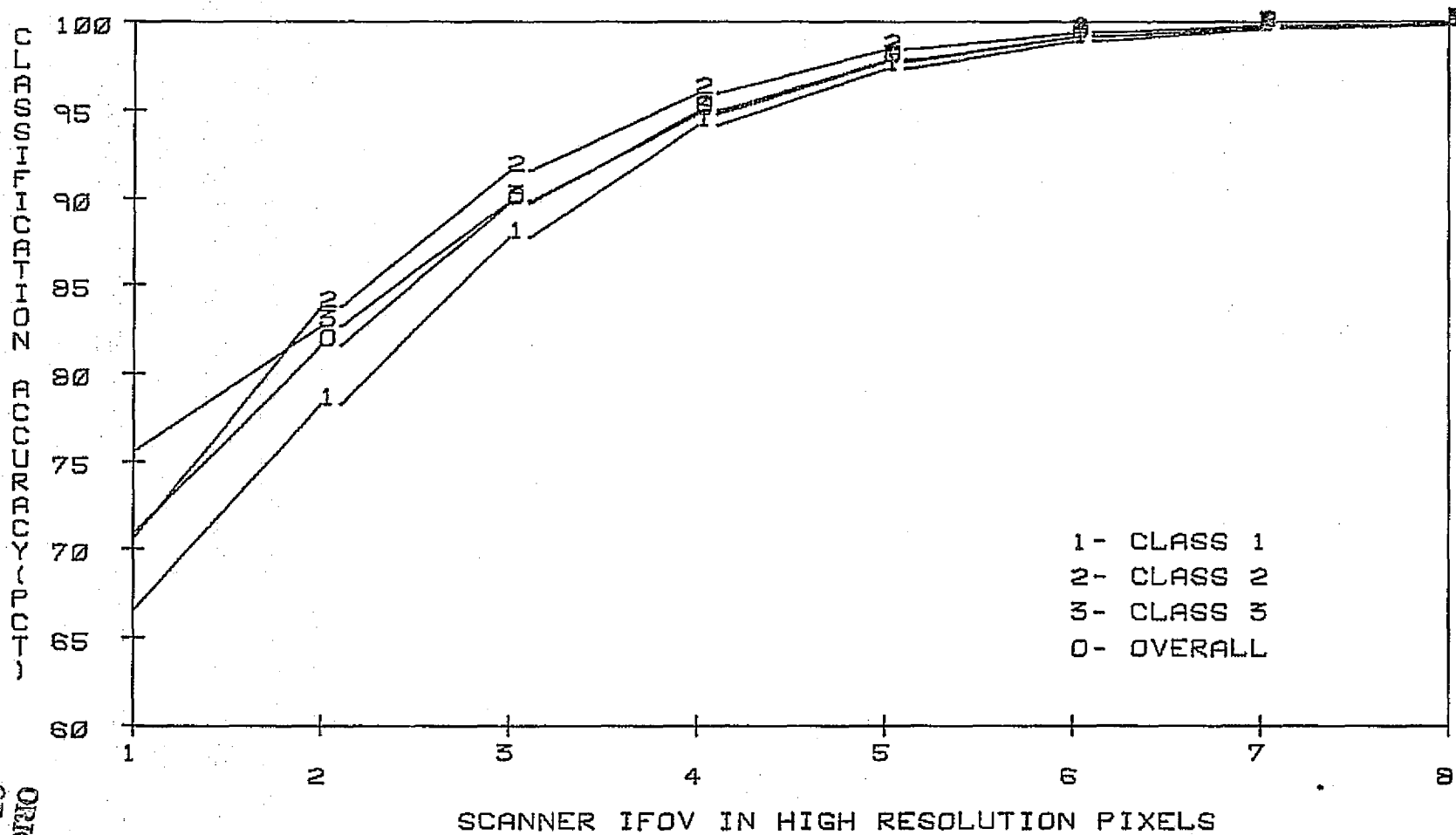


FIG. 5-6 SCANNER OUTPUT CLASSIFICATION ACCURACY VS. IFOV
ADJACENT SAMPLE CORRELATION = .55

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 5- 7 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 ADJACENT SAMPLE CORRELATION=0.60

IFOV	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{P}_c$
1	64.4	68.5	74.7	69.2
2	75.3	81.2	80.2	78.9
3	84.2	87.7	86.8	86.2
4	91.2	93.7	92.3	92.4
5	95.2	96.8	95.8	95.9
6	97.5	98.6	98.0	98.0
7	98.9	99.4	99.1	99.1
8	99.5	99.7	99.6	99.6

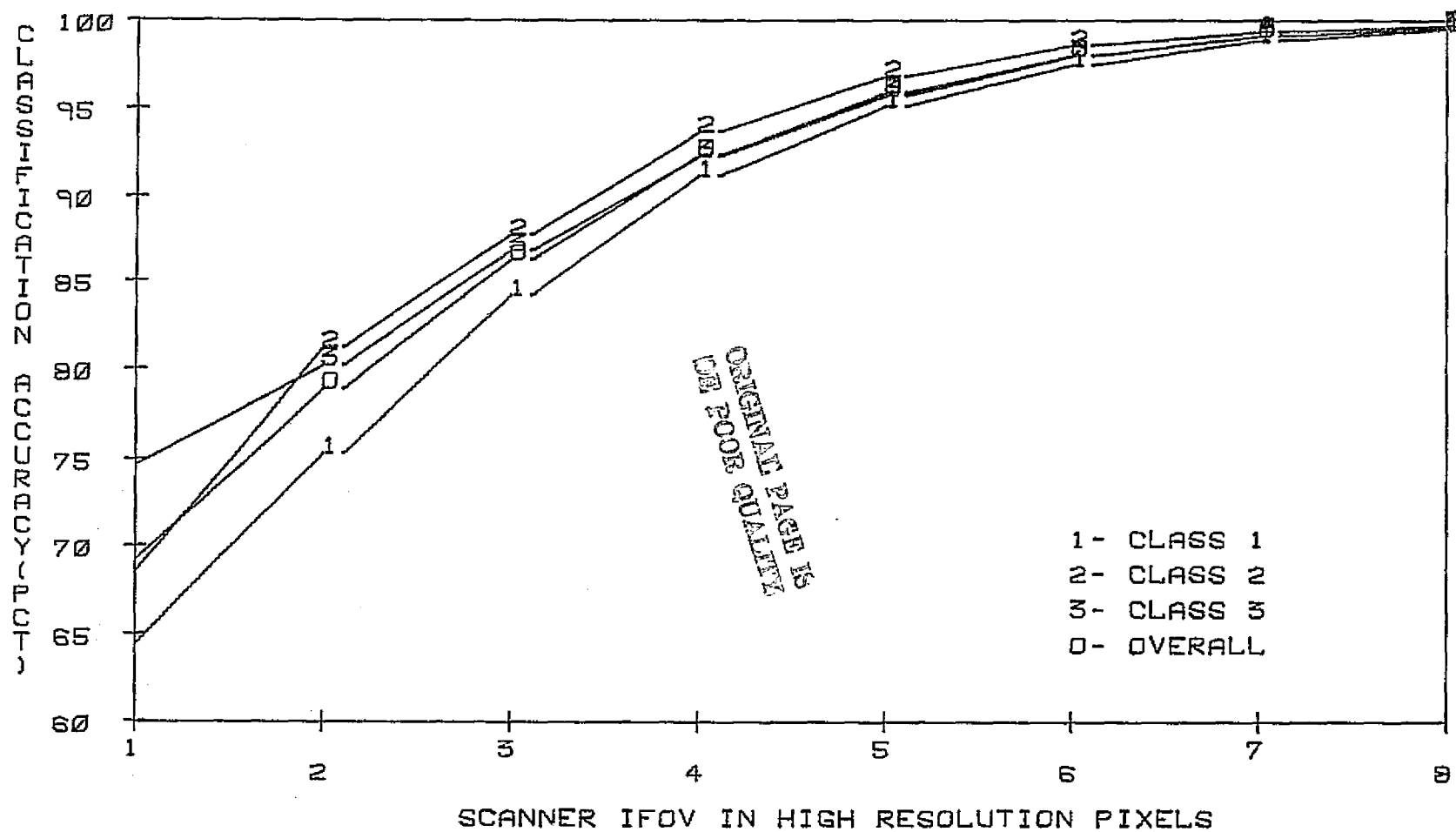


FIG. 5-7 SCANNER OUTPUT CLASSIFICATION ACCURACY VS. IFOV
ADJACENT SAMPLE CORRELATION = .6

TABLE 5- 8 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 ADJACENT SAMPLE CORRELATION=0.65

IFOV	$\hat{P}_C _{\omega_1}$	$\hat{P}_C _{\omega_2}$	$\hat{P}_C _{\omega_3}$	$\hat{P}_C$
1	63.1	66.8	73.5	67.8
2	72.3	78.6	77.9	76.3
3	80.5	84.8	83.8	83.0
4	86.7	90.4	89.0	88.7
5	92.1	94.6	93.1	93.3
6	95.4	96.9	95.9	96.1
7	97.5	98.5	97.7	97.9
8	98.7	99.2	98.9	98.9

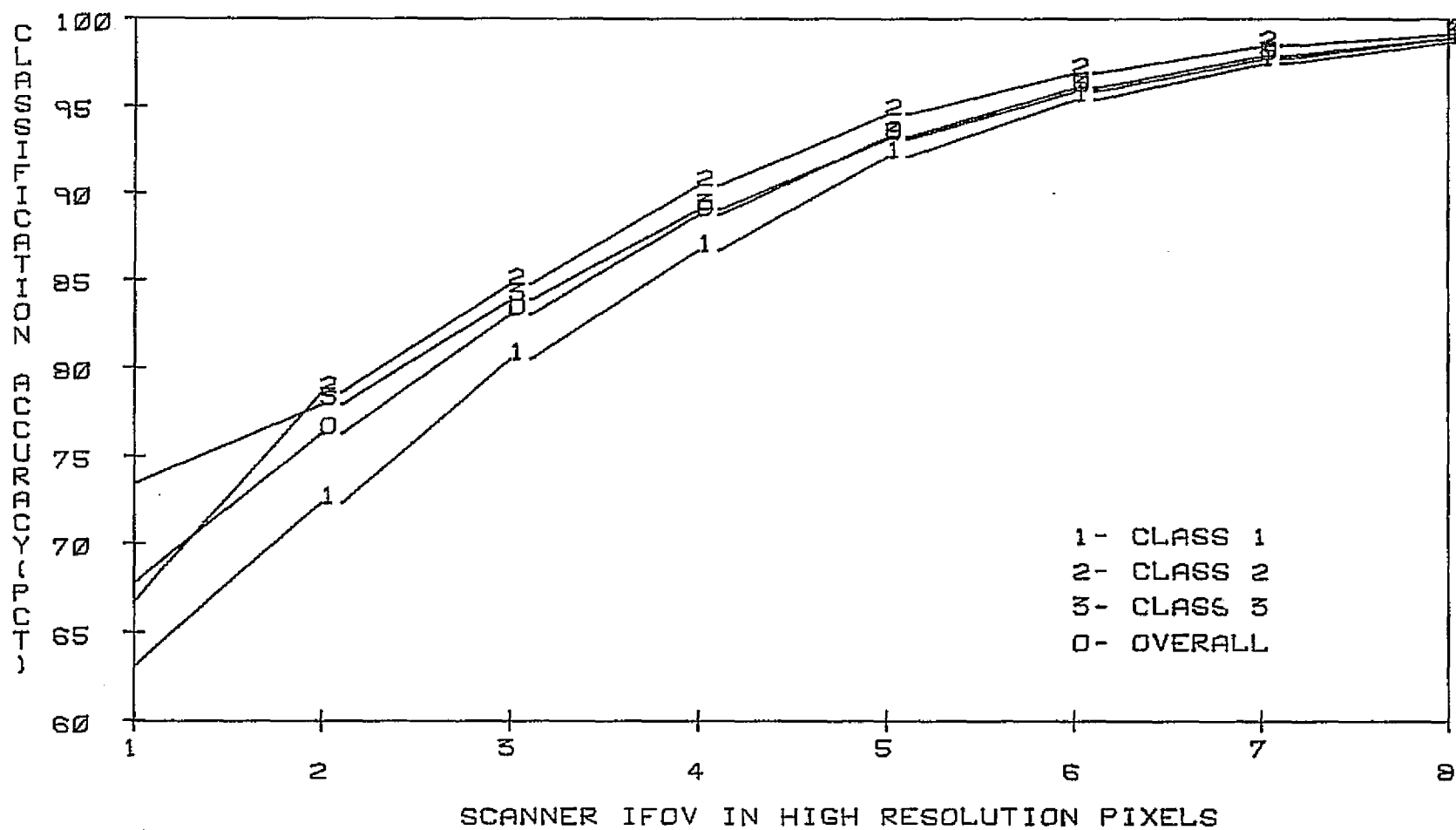


FIG. 5-8 SCANNER OUTPUT CLASSIFICATION ACCURACY VS. IFOV
ADJACENT SAMPLE CORRELATION= .65

TABLE 5- 9 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 ADJACENT SAMPLE CORRELATION=0.70

IFOV	$\hat{p}_{c \omega_1}$	$\hat{p}_{c \omega_2}$	$\hat{p}_{c \omega_3}$	$\hat{p}_c$
1	61.5	65.6	73.6	66.9
2	69.4	75.0	76.6	73.7
3	75.5	82.9	81.2	79.9
4	83.0	86.7	85.7	85.1
5	87.6	91.6	89.8	89.7
6	91.9	94.6	92.9	93.1
7	94.6	96.2	95.5	95.4
8	96.7	98.1	97.2	97.3

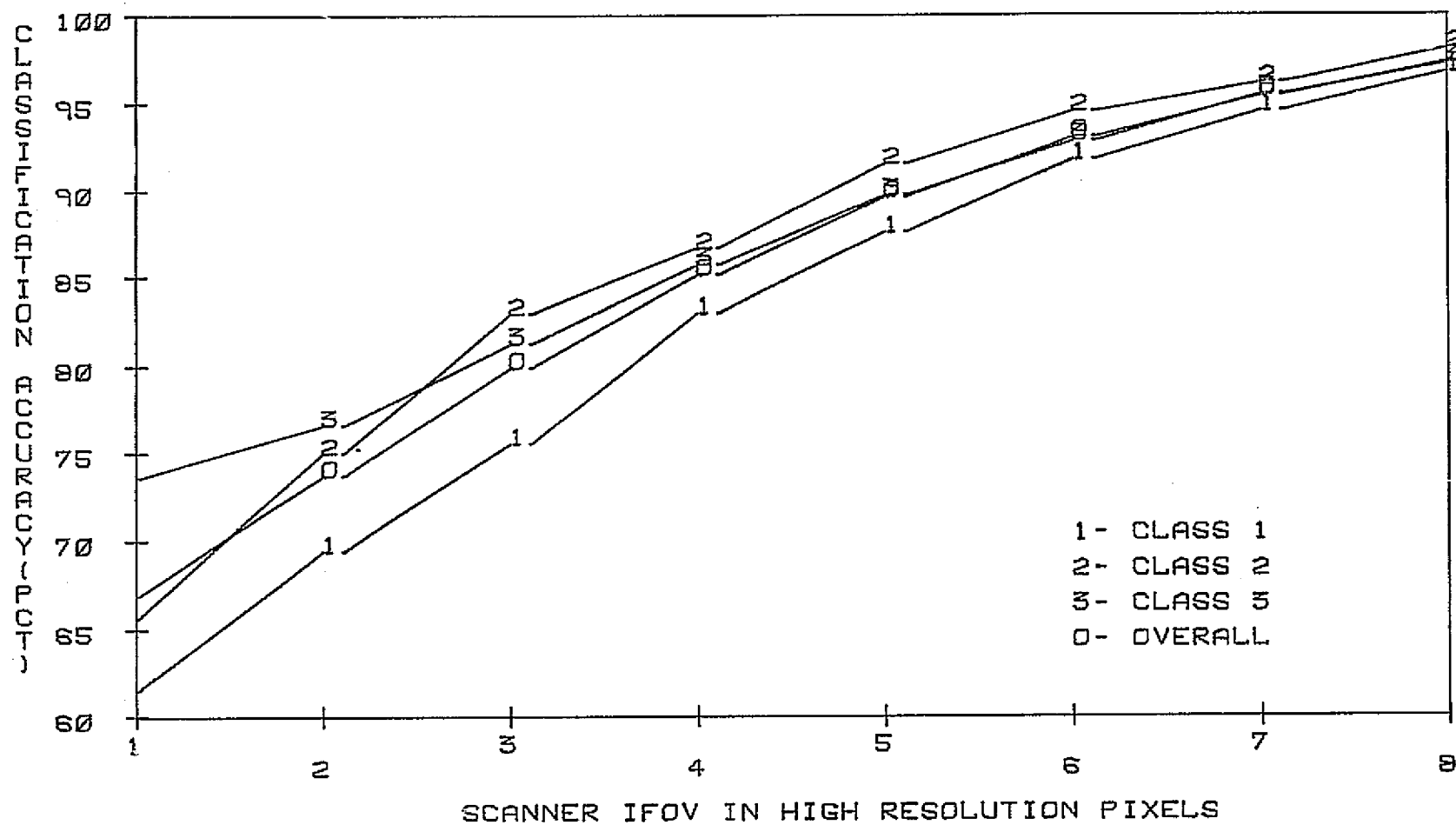


FIG. 5-9 SCANNER OUTPUT CLASSIFICATION ACCURACY VS. IFOV
ADJACENT SAMPLE CORRELATION* .7

TABLE 5-10 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
ADJACENT SAMPLE CORRELATION=0.75

IFOV	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{P}_c$
1	59.5	64.2	73.2	65.7
2	66.4	69.9	75.7	70.7
3	72.3	78.6	77.9	76.3
4	78.0	83.2	81.7	81.0
5	83.0	86.7	85.9	85.2
6	87.2	90.4	89.2	89.0
7	91.0	93.6	92.2	92.2
8	93.3	95.3	94.2	94.3

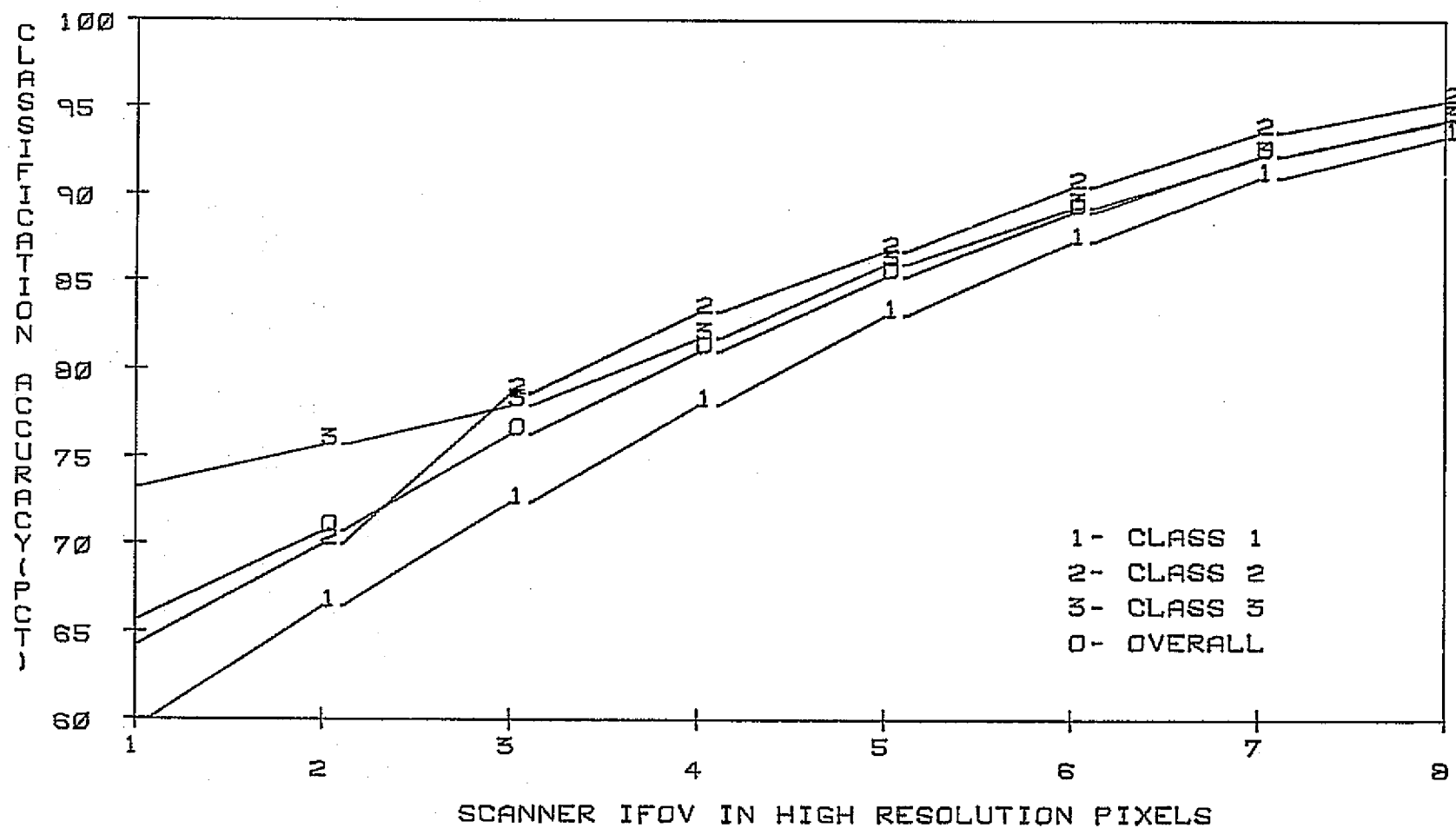


FIG. 5-10 SCANNER OUTPUT CLASSIFICATION ACCURACY VS. IFOV
ADJACENT SAMPLE CORRELATION= .75

TABLE 5-11 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
ADJACENT SAMPLE CORRELATION=0.80

IFOV	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{P}_c$
1	58.2	63.1	72.5	64.6
2	63.4	66.8	73.5	67.9
3	68.5	73.3	76.6	72.8
4	72.4	78.6	78.7	76.6
5	77.9	83.0	81.4	80.8
6	82.4	85.6	84.7	84.2
7	84.9	88.2	87.1	86.7
8	87.6	91.6	89.8	89.7

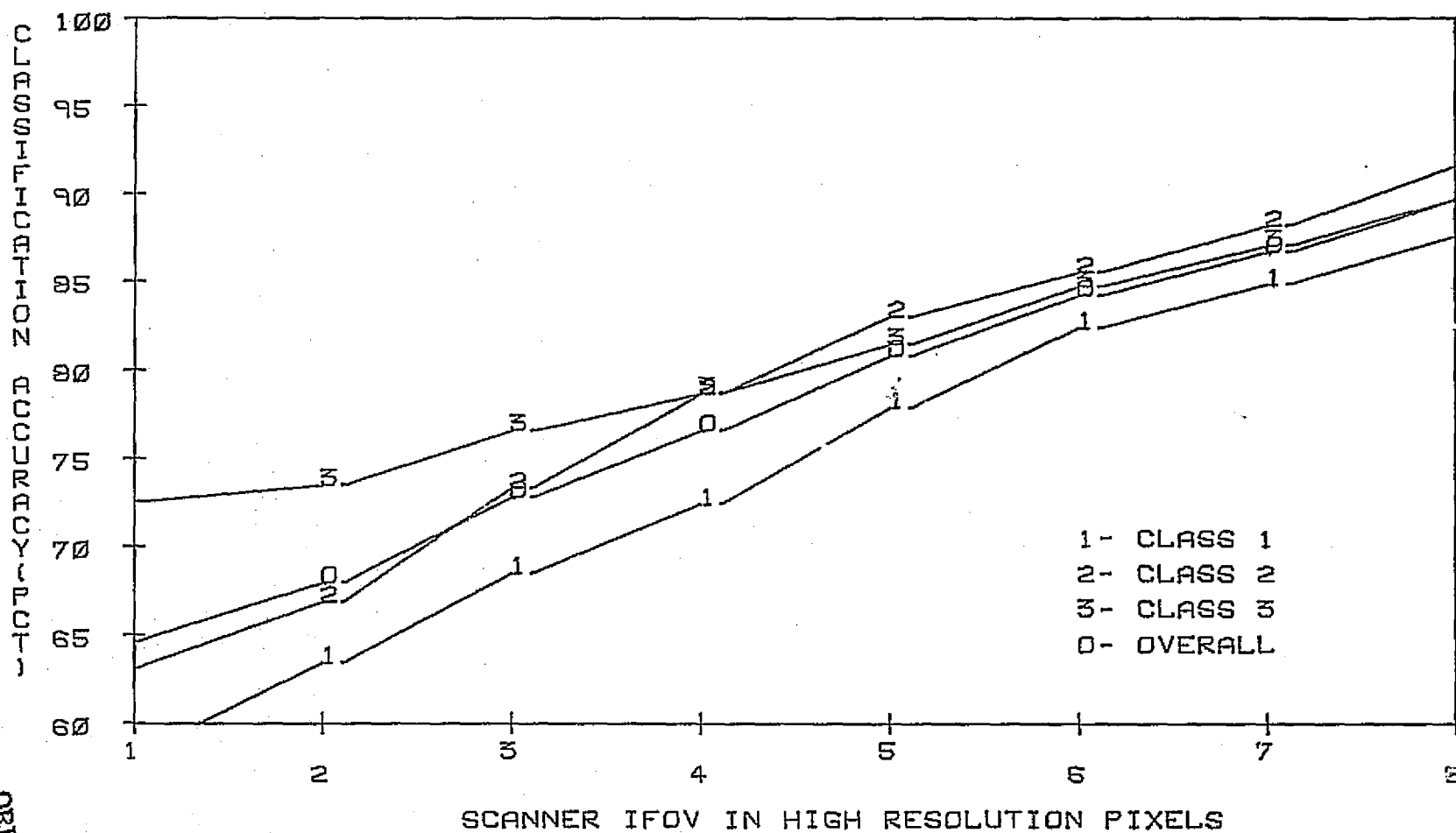


FIG. 5-11 SCANNER OUTPUT CLASSIFICATION ACCURACY VS. IFOV
ADJACENT SAMPLE CORRELATION= .8

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 5-12 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 ADJACENT SAMPLE CORRELATION=0.85

IFOV	$\hat{P}_C w_1$	$\hat{P}_C w_2$	$\hat{P}_C w_3$	$\hat{P}_C$
1	56.8	62.2	72.5	63.8
2	60.9	65.5	73.4	66.6
3	64.4	68.2	74.0	68.9
4	67.9	73.1	76.3	72.4
5	71.5	78.1	77.1	75.6
6	74.9	80.6	79.5	78.3
7	77.9	83.1	81.5	80.8
8	80.7	84.8	83.9	83.1

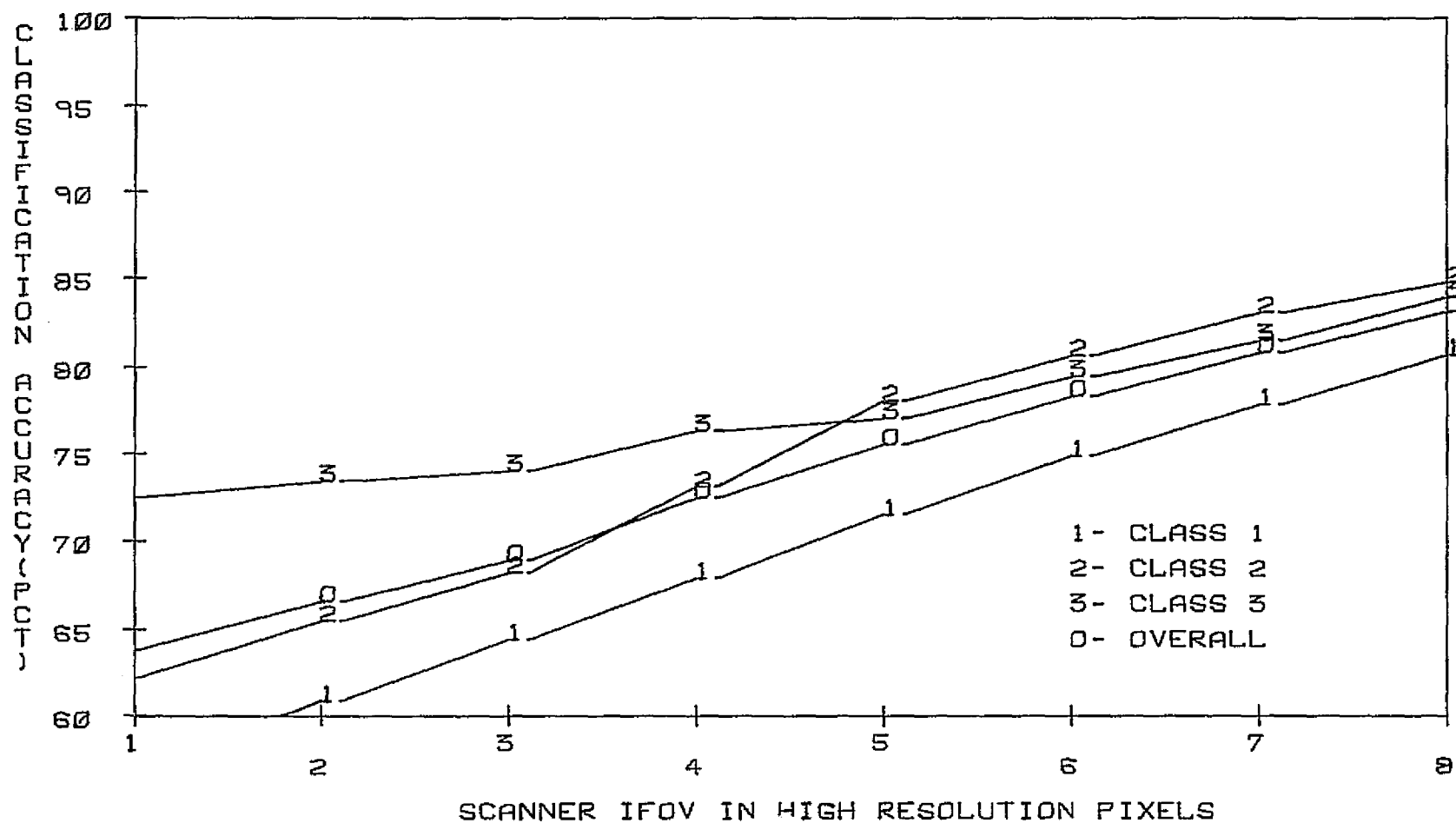


FIG. 5-12 SCANNER OUTPUT CLASSIFICATION ACCURACY VS. IFOV
ADJACENT SAMPLE CORRELATION= .95

TABLE 5-13 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 ADJACENT SAMPLE CORRELATION=0.90

IFOV	$\hat{P}_{C \omega_1}$	$\hat{P}_{C \omega_2}$	$\hat{P}_{C \omega_3}$	$\hat{P}_C$
1	55.0	61.2	73.0	63.1
2	58.2	63.1	72.7	64.7
3	60.8	65.0	73.3	66.4
4	61.9	66.5	73.5	67.3
5	65.4	68.9	74.7	69.7
6	67.7	71.1	76.3	71.7
7	69.4	75.0	76.8	73.8
8	72.3	78.6	77.8	76.2

ORIGINAL PAGE IS
 OF POOR QUALITY

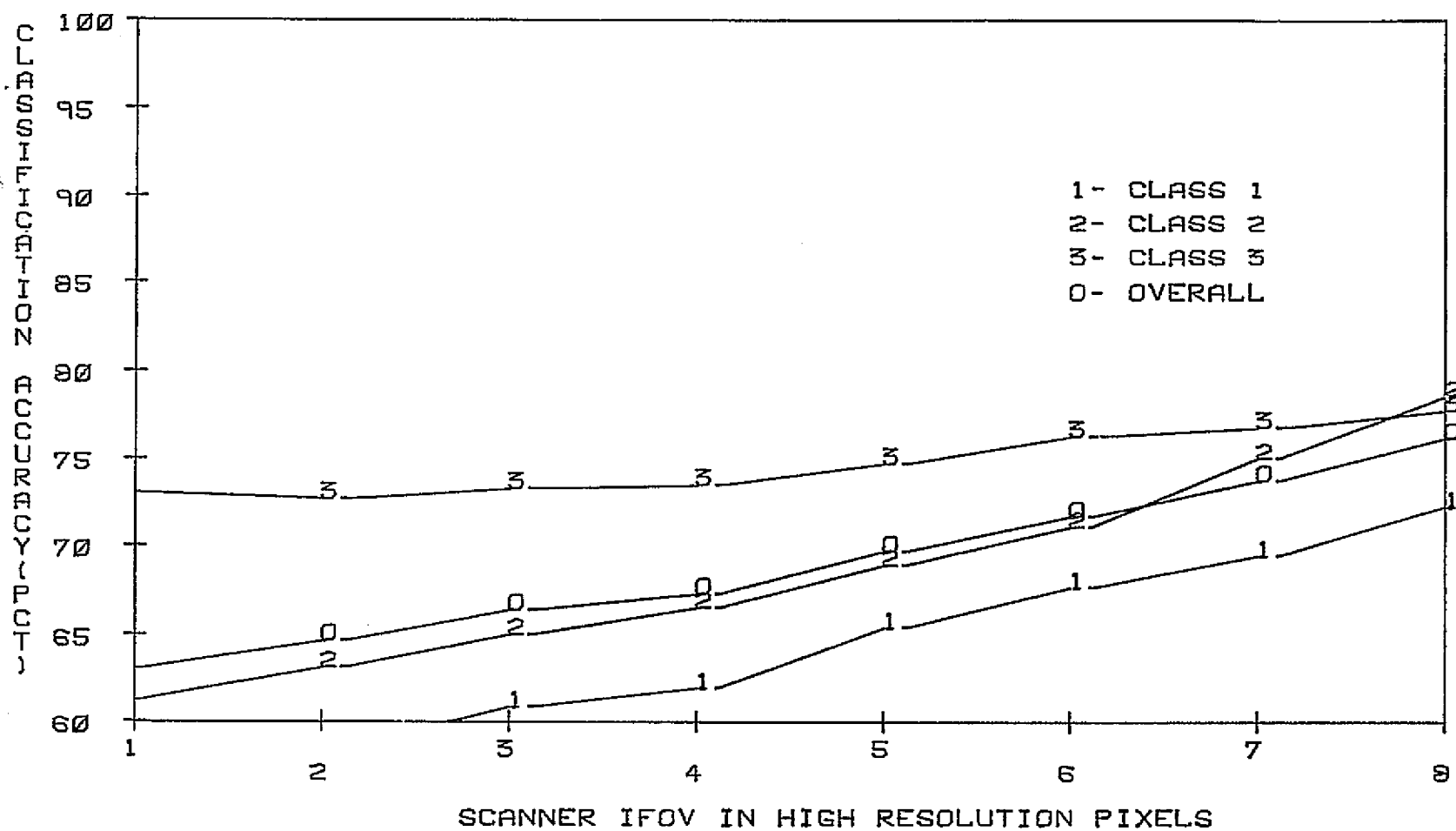


FIG. 5-13 SCANNER OUTPUT CLASSIFICATION ACCURACY VS. IFOV
ADJACENT SAMPLE CORRELATION = .9

TABLE 5-14 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 ADJACENT SAMPLE CORRELATION=0.95

IFOV	$\hat{P}_{C \omega_1}$	$\hat{P}_{C \omega_2}$	$\hat{P}_{C \omega_3}$	$\hat{P}_C$
1	54.1	60.3	72.1	62.1
2	55.0	61.0	73.0	63.0
3	56.2	62.2	72.5	63.6
4	57.9	62.9	72.7	64.5
5	58.7	63.3	72.9	65.0
6	59.7	64.4	73.3	65.8
7	61.5	65.6	73.6	66.9
8	61.9	66.2	73.5	67.2

ORIGINAL PAGE IS
 OF POOR QUALITY

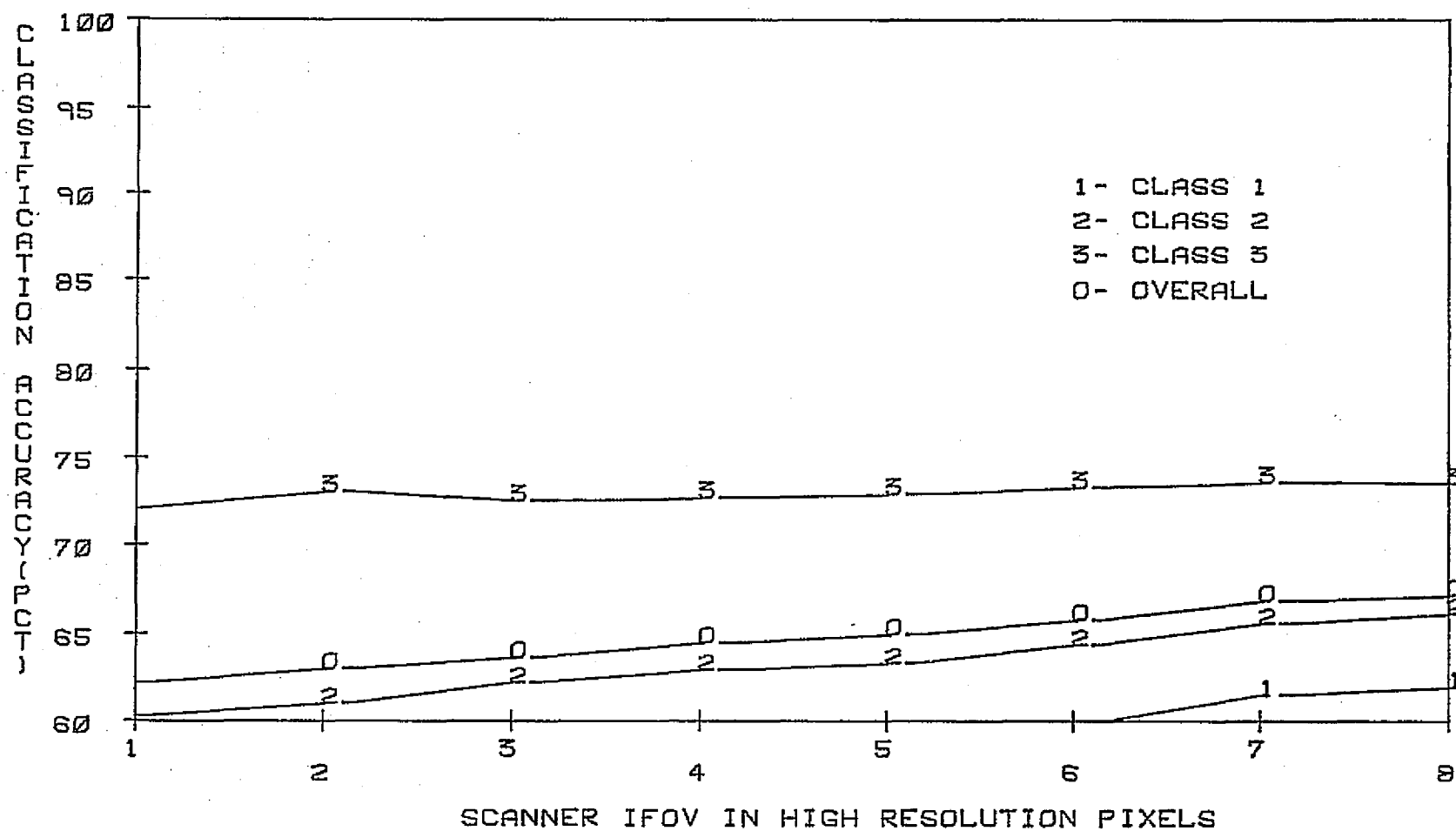


FIG. 5-14 SCANNER OUTPUT CLASSIFICATION ACCURACY VS. IFOV
ADJACENT SAMPLE CORRELATION = .95

The second property universally observed is the exponential type rise of $\hat{P}_{c|\omega_i}$ precipitated by the changing slope of the curves for a fixed ρ_x^τ and ρ_y^η . This property is brought about by the nonlinear weighting feature of W_s as the IFOV is varied. Let

$$\begin{aligned}\Delta_1 &= W(\rho_x, \rho_y, r_1) \\ \Delta_2 &= W(\rho_x, \rho_y, r_2) \\ \Delta_3 &= W(\rho_x, \rho_y, r_3)\end{aligned}\tag{5-15}$$

where r_1 , r_2 and r_3 are three different IFOV's increasing order. Then,

$$\Delta_3 - \Delta_2 < \Delta_2 - \Delta_1\tag{5-16}$$

Therefore, the classification accuracy improvement must necessarily taper off as IFOV increases. This last property is probably best demonstrated in Fig. 5-14 where the input process has a high degree of spatial correlation. The plots of $\hat{P}_{c|\omega_i}$ vs. IFOV are nearly flat with an overall classification improvement of 5.1%. This compares with 13.2% for $\rho_x = 0.9$, 25.1% for $\rho_x = 0.8$ and 26.7% for $\rho_x = 0.5$. For a degenerate case where $\rho_x = \rho_y = 1$, the characteristic function indicates that input and output classification accuracies are identical. This of course is predictable since total spatial correlation is tantamount to a process with only a DC value.

ORIGINAL PAGE IS
OF POOR QUALITY

5.3.2 Classification Accuracies at the MSS Output: Additive Gaussian Noise

In this subsection the definitions and conventions adopted in sec. 3.3 will be adhered to throughout. In order to study the effect of additive random noise in the classifiability of remotely sensed data, the scanner output class conditional statistics undergo the following linear transformation

$$\underline{\Sigma}_G' = \underline{\Sigma}_G + \underline{\Sigma}_N \quad (5-17)$$

where $\underline{\Sigma}_G$ is the noise free output statistics and $\underline{\Sigma}_N$ is the covariance matrix of a white noise process and as such it is also diagonal. The SNR in this case is defined on a class conditioned basis. However, the classes in the test case all have equal channel variances with equal spatial correlation parameters, therefore, the class conditional SNR is identical for all three populations. A fixed spatial correlation model with $\rho_x = 0.85$ and $\rho_y = 0.79$ is chosen and the output probability of correct classification vs. IFOV is estimated for SNR = 10, 20 and 30 dB. The noise enters the system at the MSS output and models the quantization and detector noise. Fig. 5-15 through 5-17 and Tables 5-15 through 5-17 show the interaction of noise and scanner IFOV and their effects on the output classification accuracy, Fig. 5-18 shows the dependence of $\hat{P}_c$ on IFOV with SNR as a running parameter.

Note that a fixed output SNR implies a variable noise power environment.

The functional variation of the classification accuracies vs. IFOV is essentially identical for different noise levels. $\hat{P}_{c|\omega_i}$ increases monotonically with increasing IFOV for a fixed SNR. Compared to the noise free case of Fig. 5-12, the slopes of $\hat{P}_{c|\omega_i}$ in the noise added case are relatively close. The classification accuracies, $\hat{P}_{c|\omega_1}$, $\hat{P}_{c|\omega_2}$, $\hat{P}_{c|\omega_3}$ and $\hat{P}_c$ increased 23.9%, 22.6%, 11.4% and 19.3% respectively where the corresponding numbers for SNR = 10 dB are 20.7%, 19.9%, 14.3% and 18.3% as IFOV ranged from 1 to 8 pixels. The percent improvement of the output classification accuracy vs. IFOV therefore is not heavily dependent on the output SNR in this case. The deterioration of the classification accuracies as noise power is increased is greater for larger scanner IFOV's. This is due to the fact that the coarse resolution output with a smaller variance is more susceptible to random disturbances than a process that already has an appreciable variance. This property is illustrated in Fig. 5-18 where the SNR = 10 dB curve diverges from the rest of the plots for higher IFOV's.

The tradeoff between the SNR and IFOV is also illustrated in Fig. 5-18 by observing that $\hat{P}_c$ is multiple valued, i.e. the combination of SNR and IFOV that result in a particular $\hat{P}_c$ is not unique. In the case under study

TABLE 5-15 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 SNR=10 DB ADJACENT SAMPLE CORRELATION=0.85

IFOV	$\hat{P}_C _{\omega_1}$	$\hat{P}_C _{\omega_2}$	$\hat{P}_C _{\omega_3}$	$\hat{P}_C$
1	53.5	59.3	57.8	56.9
2	56.1	64.6	59.4	60.0
3	58.7	66.3	60.3	61.8
4	60.9	68.3	62.4	63.9
5	66.8	71.2	64.5	67.5
6	70.5	74.1	65.6	70.1
7	72.5	76.8	70.0	73.1
8	74.2	79.2	72.2	75.2

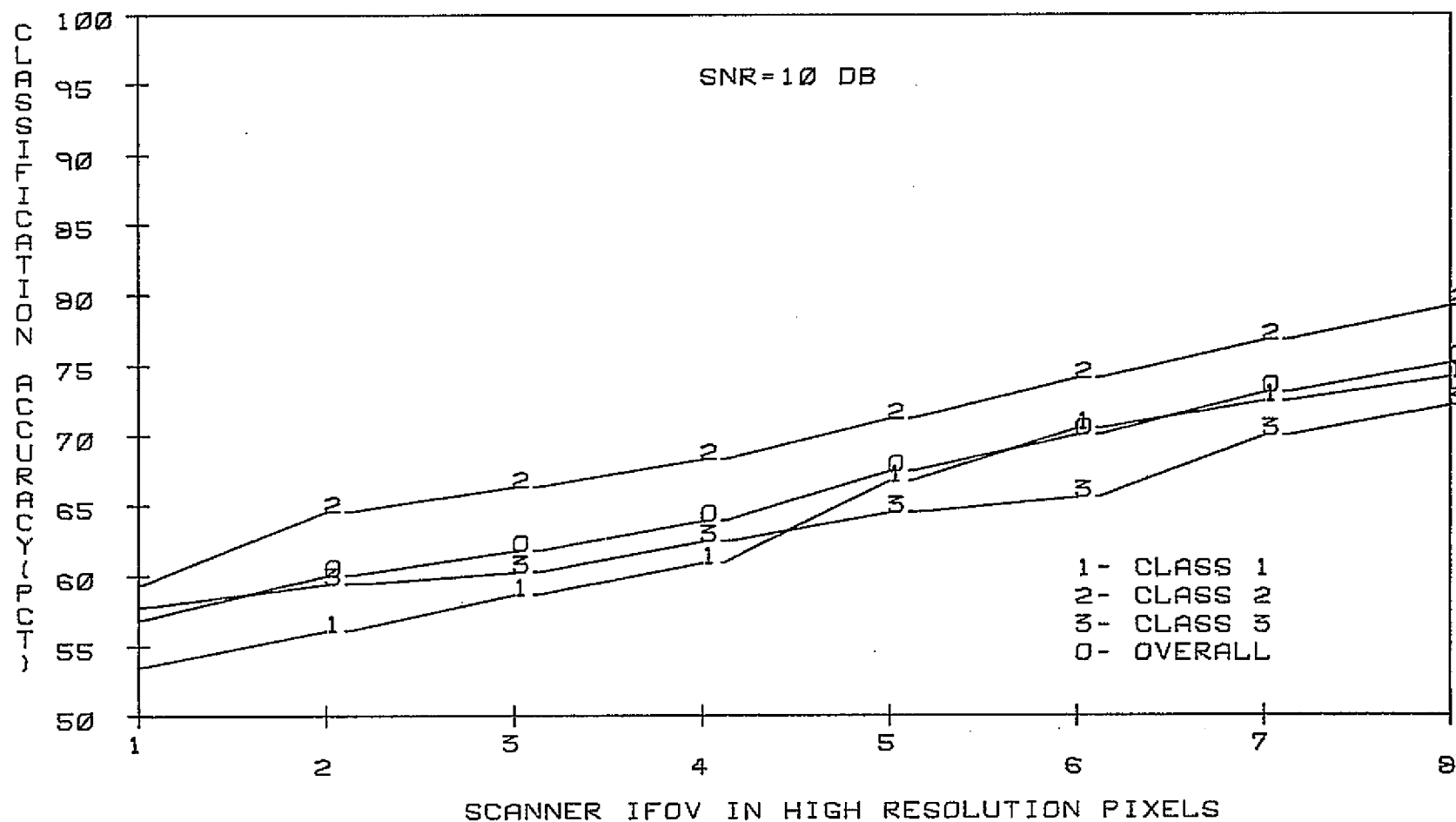


FIG. 5-15 CLASSIFICATION ACCURACIES AT THE SCANNER OUTPUT VS. ADDITIVE GAUSSIAN NOISE AND IFOV

TABLE 5-16 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 SNR=20 DB ADJACENT SAMPLE CORRELATION=0.85

IFOV	$\hat{P}_{c \omega_1}$	$\hat{P}_{c \omega_2}$	$\hat{P}_{c \omega_3}$	$\hat{P}_c$
1	56.1	63.0	70.9	63.4
2	58.9	65.7	71.0	65.2
3	62.1	67.7	73.3	67.7
4	66.7	70.6	75.1	70.8
5	70.9	75.1	76.6	74.2
6	74.3	80.2	78.5	77.7
7	77.0	83.3	79.7	80.0
8	79.9	84.9	82.1	82.3

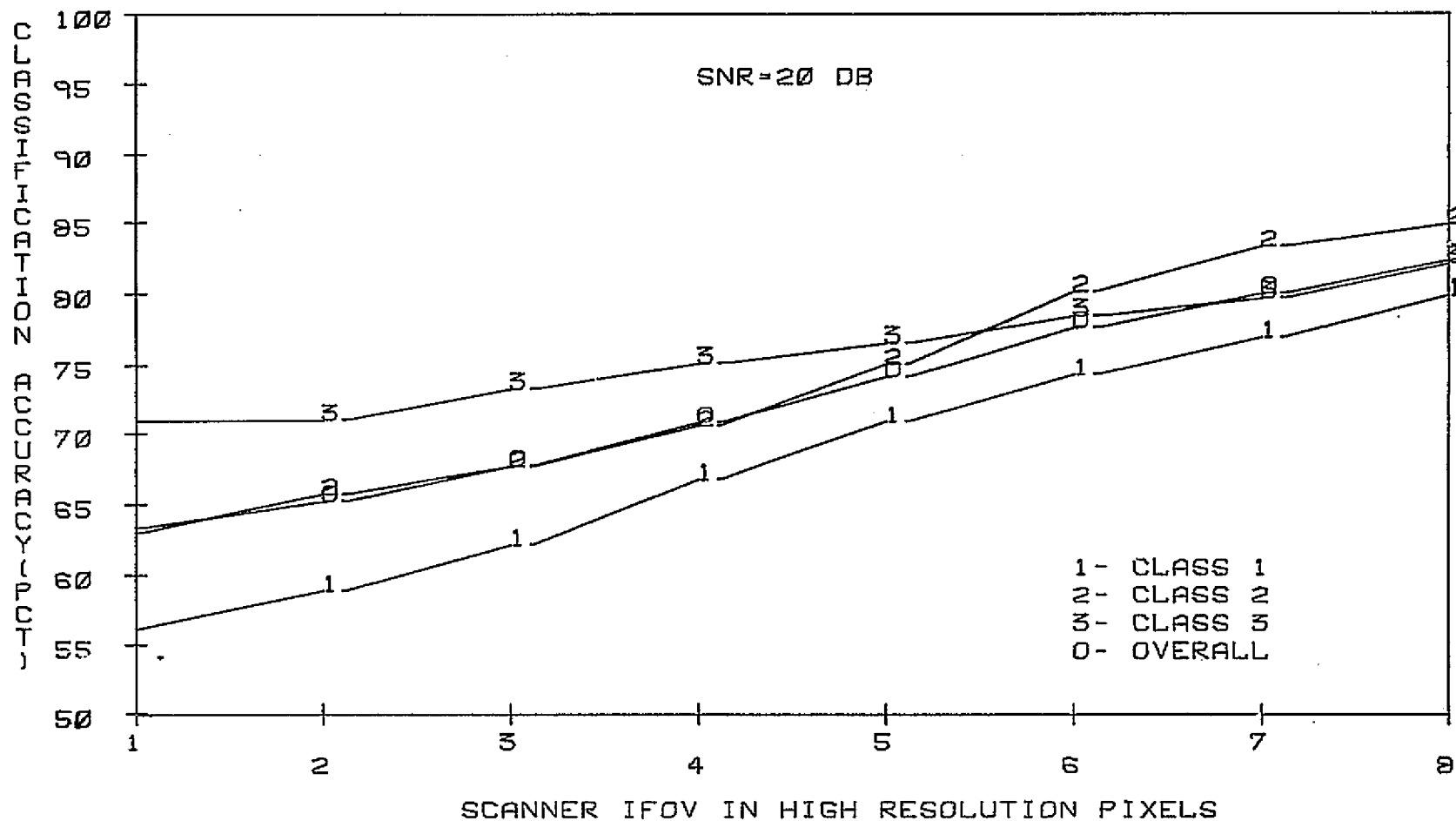


FIG. 5-16 CLASSIFICATION ACCURACIES AT THE SCANNER OUTPUT VS. ADDITIVE GAUSSIAN NOISE AND IFOV

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 5-17 SCANNER OUTPUT CLASSIFICATION ACCURACIES VS. IFOV
 SNR=30 DB ADJACENT SAMPLE CORRELATION=0.85

IFOV	$\hat{P}_C _{\omega_1}$	$\hat{P}_C _{\omega_2}$	$\hat{P}_C _{\omega_3}$	$\hat{P}_C$
1	56.2	63.5	71.7	63.8
2	60.2	65.7	73.0	66.3
3	62.9	67.9	75.5	68.8
4	68.3	70.8	76.3	71.8
5	71.5	76.2	77.8	75.2
6	74.4	80.3	79.5	78.1
7	77.1	83.5	81.6	80.7
8	80.2	85.1	83.4	82.9

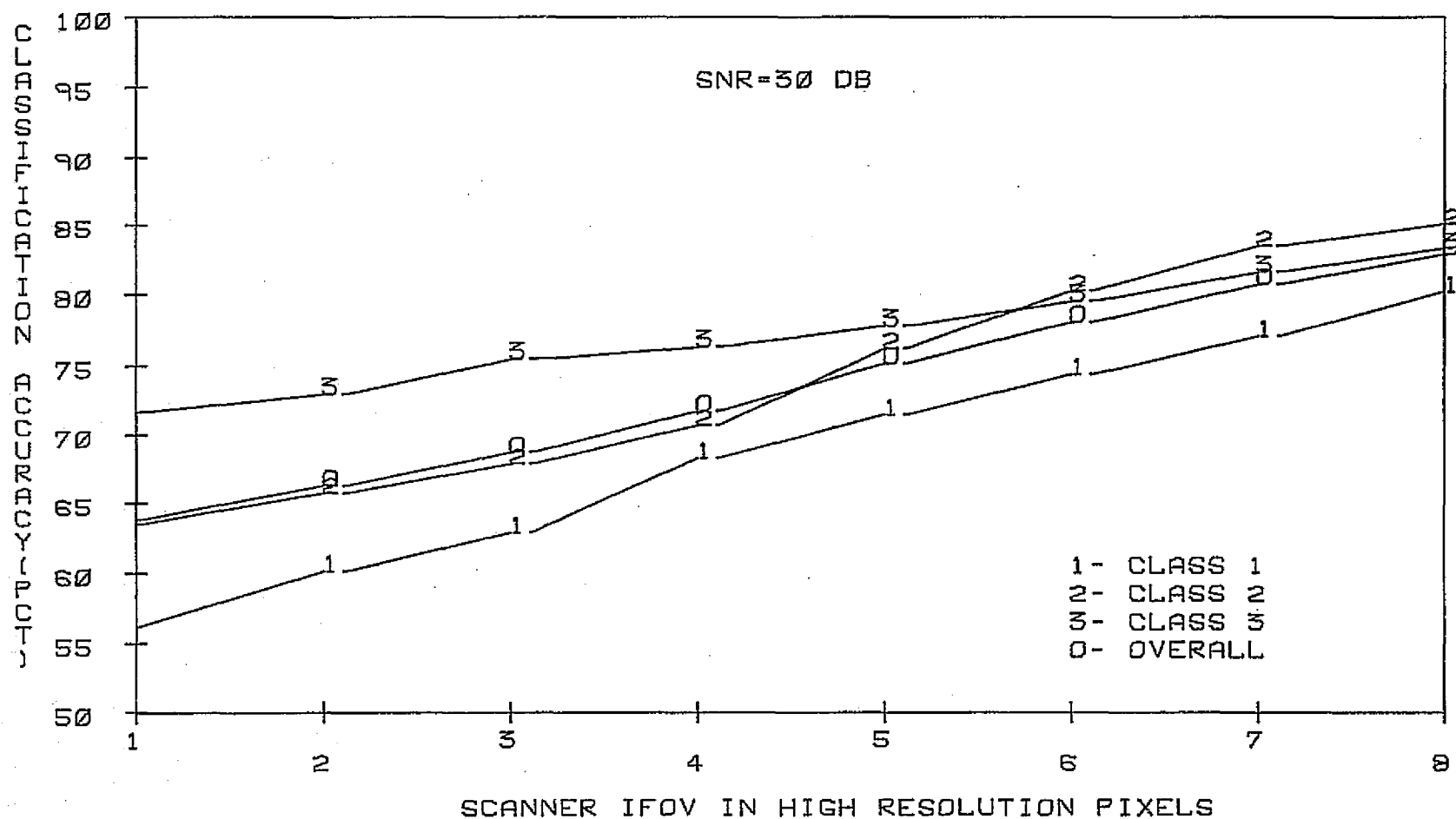


FIG. 5-17 CLASSIFICATION ACCURACIES AT THE SCANNER OUTPUT VS. ADDITIVE GAUSSIAN NOISE AND IFOV

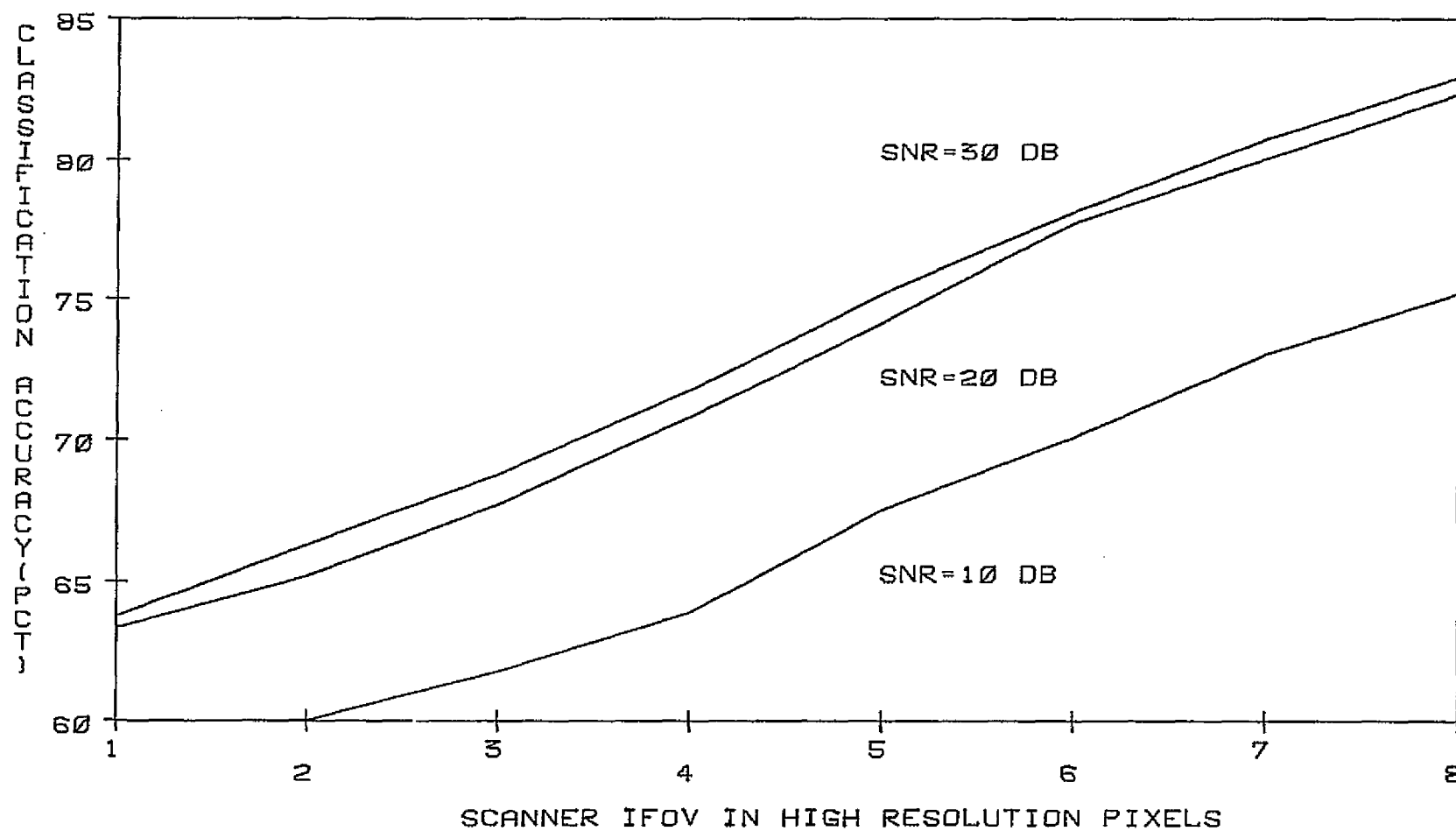


FIG. 5-18 OVERALL OUTPUT CLASSIFICATION ACCURACY VARIATION WITH NOISE AND IFOV

ORIGINAL PAGE IS
OF POOR QUALITY

a 70% classification accuracy can be achieved when IFOV = 6 pixels, SNR = 10 dB or IFOV = 3.8 pixels, SNR = 20 dB or IFOV = 3.5 pixels and SNR = 30 dB. Equivalent if the system noise level is such that the SNR is at a low 10 dB, to achieve a prescribed minimum classification performance, The resulting data spatial resolution will suffer. The same classification accuracy can be obtained with a 60% improvement in spatial resolution if the MSS is operating at a 30 dB SNR.

5.4 Summary and Conclusions

The objective of this chapter was to employ the CSP error estimation technique and MSS model in an integrated parametric package that would produce the theoretical response of the MSS in a fully controllable environment. The results presented are not intended to be exhaustive but rather to demonstrate the method and to illustrate general trends in the system response. It is constructive to compare the patterns observed with those obtained by other simulation techniques.

A parallel study aimed at the same objectives is reported in [3]. High resolution (6 m) aircraft MSS data was considered with a cascade of simulated scanner PSF's to produce data sets with 30 m, 40 m, 50 m and 60 m ground resolutions and the classification performance was estimated for each case. The results provided less than conclusive

evidence on the monotonic relationship between classification performance and the IFOV due to the very small rise in $\hat{P}_c$ as IFOV was enlarged. This conclusion can be fully understood from the theoretical curves of $\hat{P}_c$ vs. IFOV. The significant parameter, data spatial correlations, is what determines how strongly classification performance and IFOV are interrelated. As for a real data set, its spatial correlation structure is a fixed parameter. In case of high resolution aircraft data, pixel-to-pixel correlation can be as high as 0.9 or 0.95. Fig. 5-13 and 5-14 with $\rho_x = 0.9$ or 0.95 respectively clearly illustrate that $\hat{P}_c$ and IFOV are indeed weakly coupled. Had the data under investigation in [3] been less spatially correlated, this coupling would manifest itself more strongly. For satellite data having a ρ_y of about 0.75-0.8, $\hat{P}_c$ shows considerably stronger sensitivity to variations of IFOV.

The following conclusions emerge from the theoretical simulation of MSS spatial characteristics.

1. The achievable classification performance monotonically increases with increasing IFOV, at the expense of spatial resolution.
2. The degree of such dependence is directly related to the extent of spatial correlation of the random processes at the scanner input. M process with a DC value alone will have identical classification performance at the MSS input and output regardless of IFOV.

3. Additive noise, by reducing the class separabilities produces a degradation of the classification performance. For any fixed SNR, however, $\hat{P}_c$ still increases with increasing IFOV.
4. When a minimum classification performance is a design parameter, Fig. 5-18 determines the required operating states. For the test case under study, given that $\min\{\hat{P}_c\} = 70\%$, the lower bounds on IFOV are 6, 3.75 and 3 low resolution pixels for SNR, 10 dB, 20 dB and 30 dB respectively.

ORIGINAL PAGE IS
OF POOR QUALITY

CHAPTER 6

Conclusions and Observations

In this chapter we provide a broad evaluation of the results of the study and the degree which it has satisfied the objectives put forth initially. The performance of CSP Bayes error estimator was, by far, the most significant result. The transformation of ideas from abstract to practical more often than not is limited by the finiteness of the available resources; hence, it is often irrelevant whether a method is theoretically sound. In this case with the exponential rise in the number of sampling cells due to the dimensional effect, a requirement for more cells per axis would have put the usefulness of the algorithm in grave doubt. That this was not to be the case has been amply demonstrated in the experimental results of chapter 4. Admittedly the feature spaces considered cannot be classified as being of high dimensionality but within the scope of the present and near future MSS data gathered by satellites will consist of four or five bands of visible and infrared radiation. In fact even that may be reduced if some of the bands prove to be redundant in the preprocessing stages of data analysis. The systematic behavior of the estimate vs. grid size is a characteristic that

provides some degree of a posteriori information. Knowing that the estimator almost universally approaches the Bayes error with a decreasing negative bias and the fact that at about 8 cells per axis the estimate is within 1% of its final value, one may select a small grid size and choose to project the final value by heuristic or other numerical techniques. This approach may be useful when the data is of unusually high dimensionality.

There are undoubtedly a number of refinements that could accelerate the rate of convergence even further. It has been mentioned frequently that the boundary cells are the primary source of the estimation error. By adopting a larger grid, the measurement space is divided into finer partitions indiscriminantly. The optimum strategy should sample the interior of Γ_1 as coarsely as possible and the boundary region as finely as possible. One such technique is to first 'detect' the boundary by a coarse grid and then perform the partitioning by working around that segment while leaving the interior grid intact. In implementing this modification, however, close attention should be paid to the theoretical convergence property of the modified estimator. The choice of sampling grids other than binomial may be considered although one such grid with variable cell size was employed with no discernible improvement in performance.

The evaluation of the scanner spatial model provided the first application of the CSP error estimator. Compared to simulation techniques using a large data base, manipulation of the MSS parameters proved to be much simpler. The problem was simplified somewhat by the availability of closed form relationships governing the input-output statistical dependencies. This was possible because of the particular approximating function and for the scanner's PSF. The spatial characteristic function is by no means bound by such an assumption. The technique employed in Appendix A can be carried out for any specified PSF in which case the results in general are not in closed form. The observed response of the MSS was in close agreement with the reported results based on Monte-Carlo techniques. The primary difference was a far greater flexibility provided by the scanner model in examining the response to various parameter manipulations.

The number and kind of potential applications of the analysis package developed here are far greater than there was space to elaborate. The spatial model can be expanded to include a greater range of noise levels and sources. It is possible to accumulate a catalogue of system response curves corresponding to combinations of different scanner and ground scene parameters. A set of desired system parameters can be specified and the remaining set determined from the theoretical response characteristics.

The availability of several different sets of parameters to achieve a certain performance index underlines the inherent tradeoffs in designing a multispectral scanner system. The fundamental function of this parametric package, therefore, is to provide for an easily implementable technique to evaluate the system's performance with maximum flexibility and minimum input information.

ORIGINAL PAGE IS
OF POOR QUALITY

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Landgrebe et al., "Analysis Research for Earth Resources Information Systems: Where Do We Stand?" Purdue Univ/Laboratory for Application of Remote Sensing Information Note 062273, 1973
2. National Aeronautics and Space Administration Earth Resources Technology Satellite Data Users Handbook, NASA Document No. 71504249
3. Landgrebe D. A. et al., "An Empirical Study of Scanner System Parameters," Purdue Univ/Laboratory for Applications of Remote Sensing Information Note 110976, 1976
4. Crane, R. B., Malila, W. A. and Richardson, W., "Suitability of the Normal Density Assumption for Processing Multispectral Scanner Data," IEEE Trans. Geoscience Electronics, Vol. GE-10, No. 4, Oct. 1972
5. Som, R. K., A Manual of Sampling Techniques, Heinemann, 1973.
6. Anderson, T. W., Das Gupta, S., Styan G. P. H., A Bibliography of Multivariate Statistical Analysis, Oliver & Boyd, Edinburgh, 1972
7. Pearson, K., "On the Coefficient of Racial Likeness," Biometrika, 18, 1926
8. Barnard, N. M., "The Secular Variations of Skull Characters in Four Series of Egyptian Skulls," Annals of Eugenics, 6, 1935
9. Fisher, R. A., "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, 7, 1936
10. Neyman, J., and Pearson, E. S., "On the Testing of Statistical Hypothesis in Relation to Probability a priori," Proc. Camb. Phil. Soc., 9, 1933
11. Welch, B. L., "Note on Discriminant Functions," Biometrika, 31, 1939

12. Wald, A., "On a Statistical Problem Arising in the Classification of an Individual into One of Two Groups," Ann. Math. Statist., 15, 1944
13. Von Misses, R., "On the Classification of Observation Data into Distinct Groups," Ann. Math. Statist., 16, 1945
14. Rao, C. R., Advanced Statistical Methods in Biometric Research, Wiley, New York, 1952
15. Anderson, T. W., "Classification by Multivariate Analysis," Psychometrika, 16, 1951
16. Bowker, A. H., "A Representation of Hotelling's T^2 and Anderson's Classification Statistic," Contrib. Probability and Statistics (Hotelling Vol), 1960
17. Bowker, A. H. and Sitgreaves, R., "An Asymptotic Expansion for the Distribution Functions of the W-classification statistic," Stud. Item. Anal. Pred. (H. Solomon Ed.), Stanford Univ. Press, 1961
18. Teichroew, D. and Sitgreaves, R., "Computation of an Empirical Sampling Distribution for the W-classification Statistic," Ibid.
19. Okamoto, M., "Discrimination for Variance Matrix," Osaka Math. Jour., 1961
20. John, S., "The Distribution of Wald's Classification Statistic when the Dispersion Matrix is Known," Sankhya, 21, 1959
21. John, S., "On Some Classification Statistics I, II," Sankhya, 22, 1960. (Correction: Sankhya, 23, 1961)
22. John S., "Errors in Discrimination," Ann. Math. Statist. 32, 1961
23. Dunn, O. J. and Varady, P. V., "Probabilities of Correct Classification in Discrimination Analyses," Biometrics, 22, 1966
24. Lachenbruch, P. A., "An Almost Unbiased Method of Obtaining Confidence Intervals for the Probability of Misclassification in Discriminant Analysis," Biometrics, 23, 1967
25. Hills, M., "Allocation Rules and Their Error Rates," J. Roy. Statist. Soc. Ser. B, 28, 1966

26. Lachenbruch, P. A., and Mickey, M. R., "Estimation of Error Rates in Discriminant Analysis," Technometrics, 10, 1968
27. Glick, N., "Estimating Unconditional Probabilities of Correct Classification," Stanford Univ. Dept. Statistics, Tech. Report No. 3, 1969
28. Cooper, P. W., "Hyperplanes, Hyperspheres and Hyper-quadratics as Decision Boundaries," Comput. Info. Sci. (Tau & Wilcox), 1964
29. Han, C. P., "Distribution of Discriminant Function When Covariance Matrices are Propositional," Ann. Math. Statist., 40, 1969
30. Han, C. P., "Distribution of Discriminant Function in Circular Models," Ann. Inst. Statist. Math., 22, 1970
31. Gilbert, E. S., "The Effect of Unequal Variance-Covariance Matrices on Fisher's Linear Discriminant Function," Biometrics, 25, 1969
32. Chaddha, R. L., and Marcus, L. F., "An Exmpirical Comparison of Distance Statistics for Populations with Unequal Covariance Matrices," Biometrics, 24, 1968
33. Fukunaga, K., and Krile, T. F., "Calculation of Bayes Recognition Errors for Two Multivariate Gaussian Distribution," IEEE Trans. on Computers, Vol. C-18, No. 3, 1969
34. Marill, T., and Green, D. M., "On the Effectiveness of Receptors in Recognition Systems," IEEE Trans on Information Theory, Vol. IT-9, 1963
35. Cacoullos, T., "Comparing Mahalanobis distances, I: Comparing distances between k normal populations and another unknown," Sankhya, 27, 1965
36. Lachenbruch, P. A., "Some Results on Multiple Group Discrimination Problem," Discriminant Analysis and Applications, Cacoullon Ed, Academic Press, 1973
37. Subrahmanian, K., Multivariate Analysis: A Selected and Abstracted Bibliography. 1957-1972, Marcel Dekker, Inc., 1973
38. Cacoullos, T., and Styan, G. P., "A Bibliography of Discriminant Analysis," Discriminant Analysis and Applications, Cacoullos Ed., Academic Press, 1973

39. Lachenbruch, P. A., Discriminant Analysis, 1975
40. Toussaint, G. T., "Bibliography on Estimation of Misclassification," IEEE Trans. Information Theory, Vol. IT20, No. 4, 1974
41. Fukunaga, K., Introduction to Statistical Pattern Recognition, Academic Press, 1972
42. Van Trees, H. T., Detection, Estimation and Modulation Theory: Part I, Wiley, 1968
43. Clenshaw, C. W., and Curtis, A. R., "A Method for Numerical Integration on an Automatic Computer," Numer. Math., 2, 1960
44. Cooper, P. W., "The Hypersphere in Pattern Recognition," Information and Control, 5, 1962
45. Cooper, P. W., "Statistical Classification with Quadratic Forms," Biometrika, 50, 1963
46. Cooper, P. W., "Quadratic Discriminant Functions in Pattern Recognition," IEEE Trans. Information Theory, IT-11, 1965
47. Davis, P. J., and Rabinowitz, P., Methods of Numerical Integration, Academic Press, 1975
48. Molenaar, W., Approximations to the Poisson, Binomial and Hypergeometric Distribution Functions, Amsterdam Mathematisch Centrum 1970
49. Stroud, A. H., Approximate Calculation of Multiple Integrals, Prentice-Hall, 1971
50. Holter, M. R., et al., "Imaging with Nonphotographic Sensors," Remote Sensing with Special Reference to Agriculture and Forestry, National Academy of Science, 1970
51. Goldberg, I. L., "Design Consideration for a Multi-spectral Scanner for ERTS," Symposium on Information Process, Purdue Univ., 1969
52. Bates, J. C., and Dumar, H. J., "High Efficiency Conical Scanner for Earth Resources Application," Scanners and Imagery Systems for Earth Observation, Proceeding of SPIE, 1974
53. Riemer, T. E. and McGillem, C. D., "Constrained Optimization of Image Restoration Filters," Applied Optics, 12, 1973

54. McGillem, C. D., Riemer, T. E., Mobasser, B. G.,
"Resolution Enhancement of ERTS Imagery," Proceedings
of Machine Processing of Remotely Sensed Data,
LARS, Purdue Univ., IEEE Catalog Number 75CH1009-6,
1975
55. Hunt, B. R., "Deconvolution of Linear Systems by
 Constrained Regression and its Relationship to the
 Wiener Theory," IEEE Trans. Autom. Control, 1972
56. Ready, P. J., and Wintz, P. A., "Multispectral Data
 Compression Through Transform Coding and Block
 Quantization," Purdue Univ. TR-EE 72, 1972
57. Kettig, R. L., "Computer Classification of Remotely
 Sensed Multispectral Image Data by Extraction and
 Classification of Homogenous Objects," Purdue
Univ. LARS Information Note 050975
58. Box, G. E., and Jenkins, M., Time Series Analysis:
Forecasting and Control, Holden-Day, 1976
59. Elterman, L., "Atmospheric Attenuation Model in the
 Ultraviolet, Visible and Infrared Regions for Altitude
 up to 50 Km," Res. Paper #46, Opt. Phys. Lab., Air
Force Cambridge Research Labs
60. National Aeronautics and Space Administration Earth
Resources Landsat Program Users Handbook, NASA Document
No. 765054258
61. Billingsley, F. C., "Noise Consideration in Digital
 Image Processing Hardware," Picture Processing and
Digital Filtering, Ed. T. S. Huang, Springer-Verlag,
 New York, Heidelberg, Berlin, 1975
62. Robinson, E. A., Statistical Communication and Detec-
tion, London, Charles Griffin and Co., 1967
63. Friedman, H. D., "On the Expected Error in the
 Probability of Misclassification," Proc. IEEE, April
 1965
64. Ready, P. J., Wintz, P. A., Whitsitt, S. J., and
 Landgrebe, D. A., "Effects of Compression and Random
 Noise on Multispectral Data," Proc. 7th Symposium on
Remote Sensing of the Environment, Univ. of Michigan,
 1971

ORIGINAL PAGE IS
 OF POOR QUALITY

65. Landgrebe, D. A. et al., Final Report, Laboratory for Applications of Remote Sensing NASA Contract NAS9-14016
June 1, 1975 - May 31, 1976
66. Papoulis, A., Probability, Random Variables and Stochastic Processes, McGraw Hill

APPENDICES

Appendix A

Multispectral Scanner Output Statistics

In order to determine the effects of different scanner IFOV's and their interaction with classification accuracy of a data set, it is essential that the required output covariance matrices be parametrically represented in terms of known input quantities. In sec. 3.2.2 it was noted that the entire spectral covariance matrix is specified if the appropriate spatial correlation functions are known. Let $f(x,y)$, $g(x,y)$ and $h(x,y)$ denote the input and output random processes associated with any two matching bands and the scanner PSF, respectively. It is well known that the above quantities are related by a convolution integral.

$$g(x,y) = \iint f(x-\lambda_1, y-\lambda_2) h(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \quad (A-1)$$

In order to derive specific results, two different scanner PSF's are considered: (a) a spherically symmetric Gaussian PSF; and (b) a rectangular PSF. The spatial correlation matrix describing the scene is a two sided exponential.

A.1 Gaussian Scanner PSF

The PSF and spatial correlation model are given by

$$\begin{aligned} R_{ff}(\tau, \eta) &= \rho_o^{|\tau|} \rho_o^{|\eta|} \\ h(x, y) &= c_1 e^{-\frac{x^2}{r_o^2}} e^{-\frac{y^2}{r_o^2}} \end{aligned} \quad (A-2)$$

where $\rho_o = e^{-a}$ is the adjacent pixel correlation assumed equal along the horizontal and vertical directions. This assumption is not in contradiction with the fact that in a digital data set sample-to-sample correlation is higher than line-to-line correlation because of the closer physical distance between the samples. In continuous domain, such as this formulation, where theoretically equally spaced lines and columns can exist, there is little reason for assuming different pixel-to-pixel correlation along each direction. Two quantities, c_1 and r_o specify the PSF where c_1 is a normalizing constant providing unity gain and r_o is the filter's characteristic length, closely related to the IFOV.

With the parameters of the problem defined, the scanner output correlation function can be expressed as;

$$S_{gg}(u, v) = S_{ff}(u, v) |H(u, v)|^2 \quad (A-3)$$

where $S(u, v)$ is spectral density. Let $M(u, v) = |H(u, v)|^2$, then

ORIGINAL PAGE IS
OF POOR QUALITY

$$R_{gg}(\tau, \eta) = R_{ff}(\tau, \eta) * m(\tau, \eta) \quad (A-4)$$

$$m(\tau, \eta) = \frac{\pi c_1^2 r_o^2}{2} e^{-\frac{\tau^2}{2r_o^2}} e^{-\frac{\eta^2}{2r_o^2}} \quad (A-5)$$

Using the separability property of the functions involved,

$$\begin{aligned} R_{gg}(\tau, \eta) &= \int R_{ff}(\tau-x) m(x) dx \int R_{ff}(\eta-y) m(y) dy \\ &= Ax B \end{aligned} \quad (A-6)$$

$$\begin{aligned} A &= \frac{\sqrt{\pi} c_1 r_o}{\sqrt{2}} \int_{-\infty}^{\infty} e^{-a|\tau-x|} e^{-\frac{x^2}{2r_o^2}} dx \\ &= \frac{\sqrt{\pi} c_1 r_o}{\sqrt{2}} \int_{-\infty}^{\tau} e^{-a(\tau-x)} e^{-\frac{x^2}{2r_o^2}} dx + \end{aligned} \quad (A-7)$$

$$\frac{\sqrt{\pi} c_1 r_o}{\sqrt{2}} \int_{\tau}^{\infty} e^{a(\tau-x)} e^{-\frac{x^2}{2r_o^2}} dx$$

Combining the exponentials and completing the squares,

$$\begin{aligned} A &= \frac{\sqrt{\pi} c_1 r_o}{\sqrt{2}} \left[e^{\frac{a^2 r_o^2}{2} - a\tau} \int_{-\infty}^{\tau} e^{-\frac{(x-ar_o^2)^2}{2r_o^2}} dx + \right. \\ &\quad \left. e^{\frac{a^2 r_o^2}{2} + a\tau} \int_{\tau}^{\infty} e^{-\frac{(x+ar_o^2)^2}{2r_o^2}} dx \right] \quad (A-8) \end{aligned}$$

The individual integrals can now be represented in terms of the Q function

$$\begin{aligned}
 A &= \pi c_1 r_o^2 \left[e^{\frac{a^2 r_o^2}{2} - a\tau} \int_{-\infty}^{\frac{\tau - ar_o^2}{r_o}} e^{-\frac{x^2}{2}} dx + \right. \\
 &\quad \left. e^{\frac{a^2 r_o^2}{2} + a\tau} \int_{\frac{\tau + ar_o^2}{r_o}}^{\infty} e^{-\frac{x^2}{2}} dx \right] \\
 &= \pi c_1 r_o^2 \left[e^{\frac{a^2 r_o^2}{2} - a\tau} Q\left(\frac{ar_o^2 - \tau}{r_o}\right) + e^{\frac{a^2 r_o^2}{2} + a\tau} Q\left(\frac{ar_o^2 + \tau}{r_o}\right) \right]
 \end{aligned}
 \tag{A-9}$$

The constant c_1 is the solution to the following equation

$$\left[\int_{-\infty}^{\infty} \sqrt{c_1} e^{-\frac{x^2}{r_o^2}} dx \right] \left[\int_{-\infty}^{\infty} \sqrt{c_1} e^{-\frac{y^2}{r_o^2}} dy \right] = 1$$

Therefore,

$$c_1 = \frac{1}{\pi r_o^2}$$

Noting that B is similarly evaluated, the spatial correlation function at the scanner output is given by

$$R_{gg}(\tau, \eta) = \left[e^{\frac{a^2 r_o^2}{2} - a\tau} Q\left(ar_o - \frac{\tau}{r_o}\right) + e^{\frac{a^2 r_o^2}{2} + a\tau} Q\left(ar_o + \frac{\tau}{r_o}\right) \right]^2
 \tag{A-11}$$

The above relationship can be easily modified to cover the case of unequal pixel-to-pixel correlation along cross track and down track directions. If $R_{ff}(\tau, \eta)$ is given by

$$R_{ff}(\tau, \eta) = e^{-a|\tau|} e^{-b|\eta|}$$

Then it follows that

$$R_{gg}(\tau, \eta) = \left[e^{\frac{a^2 r_o^2}{2} - a\tau} Q(ar_o - \frac{\tau}{r_o}) + e^{\frac{a^2 r_o^2}{2} + a\tau} Q(ar_o + \frac{\tau}{r_o}) \right] \times$$

$$\left[e^{\frac{b^2 r_o^2}{2} - b\eta} Q(br_o - \frac{\eta}{r_o}) + e^{\frac{b^2 r_o^2}{2} + b\eta} Q(br_o + \frac{\eta}{r_o}) \right]$$

(A-12)

Note that since the input process $f(x, y)$ has a unity variance $R_{gg}(0, 0)$ is in effect a weighting by which any input variance will be multiplied to produce the corresponding output spectral variance. The right hand side of (A-9), therefore, can be considered as a weighting function associated with any multiband scanner to relate input and output statistics. Denote this function by $W_g(\tau, \eta, a, b)$.

The next item of interest is the output crosscorrelation among channels. This quantity, designated by $R_{g_i g_j}(\tau, \eta)$, is a straight forward extension of the method just described. Again assuming a Markov or exponential structure

governing the crosscorrelation function between channels

$$R_{f_i f_j}(\tau, \eta) = r_{f_i f_j} \sigma_{f_i} \sigma_{f_j} e^{-a_{ij}|\tau|} e^{-b_{ij}|\eta|} \quad (A-13)$$

where r_{ij} is the spectral crosscorrelation coefficient at the input such that $|r_{f_i f_j}| \leq 1$. a_{ij} and b_{ij} are defined similar to a and b . Since the crosscorrelation function between a pair of outputs of a linear system is related to the input cross correlation of the same pair in a form similar to (A-3) [66], i.e.

$$S_{g_i g_j}(u, v) = S_{f_i f_j}(u, v) |H(u, v)|^2 \quad (A-14)$$

the same technique used previously provides the acrossband correlation function at the MSS output.

$$R_{g_i g_j}(\tau, \eta) = r_{f_i f_j} \sigma_{f_i} \sigma_{f_j} W_s(\tau, \eta, a_{ij}, b_{ij}) \quad (A-15)$$

From (A-13), the crosscorrelation coefficient between any two channels at the scanner output is

$$r_{g_i g_j} = \frac{R_{g_i g_j}(0, 0)}{\sigma_{g_i} \sigma_{g_j}} \quad (A-16)$$

ORIGINAL PAGE IS
OF POOR QUALITY

From (A-15)

$$R_{g_i g_j}(0,0) = r_{f_i f_j} \sigma_{f_i} \sigma_{f_j} W_s(0,0, a_{ij}, b_{ij}) \quad (A-17)$$

also

$$\sigma_{g_i} = \sigma_{f_i} W_s^{\frac{1}{2}}(0,0, a_{ii}, b_{ii}) \quad (A-18)$$

hence

$$r_{g_i g_j} = \frac{W_s(0,0, a_{ij}, b_{ij})}{W_s^{\frac{1}{2}}(0,0, a_{ii}, b_{ii}) W_s^{\frac{1}{2}}(0,0, a_{jj}, b_{jj})} r_{f_i f_j} \quad (A-19)$$

Therefore, the band-to-band correlation coefficients are identical at scanner input and output provided spatial auto and crosscorrelation functions at the input are equivalent, i.e., $a_{ii} = a_{ij}$, $b_{ii} = b_{ij}$.

A.2 Rectangular Scanner PSF

A rectangular point spread function is defined here by

$$h(x,y) = \begin{cases} 1/r_o^2 & |x||y| \leq \frac{r_o}{2} \\ 0 & \text{otherwise} \end{cases} \quad (\text{A-20})$$

Similarly to (A-4);

$$R_{gg}(\tau, \eta) = R_{ff}(\tau, \eta) * m(\tau, \eta)$$

where

$$m(\tau, \eta) = \frac{1}{2} \left(1 - \frac{|\tau|}{r_o}\right) \left(1 - \frac{|\eta|}{r_o}\right) \quad |\tau|, |\eta| \leq r_o \quad (\text{A-21})$$

Empling (A-6) to (A-21)

$$\begin{aligned} A &= \frac{1}{r_o} \int_{-\infty}^{\infty} e^{-a|\tau-x|} \left(1 - \frac{|x|}{r_o}\right) dx \\ &= \frac{1}{r_o} \left[\int_{-r_o}^0 e^{-a(\tau-x)} \left(1 + \frac{x}{r_o}\right) dx + \int_0^{\tau} e^{-a(\tau-x)} \left(1 - \frac{x}{r_o}\right) dx \right. \\ &\quad \left. + \int_{\tau}^{r_o} e^{a(\tau-x)} \left(1 - \frac{x}{r_o}\right) dx \right] \end{aligned} \quad (\text{A-22})$$

Designating the three terms in the bracket as I, II and III, routine integration techniques yield the following results

ORIGINAL PAGE IS
OF POOR QUALITY

$$I = \frac{1}{ar_0} \left(1 - \frac{1 - e^{-ar_0}}{ar_0}\right) e^{-a|\tau|} \quad (A-23a)$$

$$II = \frac{1}{ar_0} \left(1 - \frac{1 - e^{-a|\tau|}}{a|\tau|}\right) e^{-a|\tau|} \quad (A-23b)$$

$$III = \frac{1}{ar_0} \left(\left(1 - \frac{1 - e^{-ar_0}}{ar_0}\right) - \left(1 - \frac{1 - e^{-a|\tau|}}{a|\tau|}\right) \right) e^{-a|\tau|} \quad (A-23c)$$

A similar expression is obtained for B by substituting η and b for τ and a in (A-23a) thru (A-23c). The scanner characteristic function, $W_s(0,0,a,b)$ is evaluated by equating $\tau = \eta = 0$ in I, II and III;

$$W_s(0,0,a,b) = \frac{2}{ar_0} \left(1 - \frac{1 - e^{-ar_0}}{ar_0}\right) \times \frac{2}{br_0} \left(1 - \frac{1 - e^{-br_0}}{br_0}\right) \quad (A-24)$$

(A-24) is plotted in sec. 3.2.2 for different values of a , b and r_0 .

Appendix B

Exponential Spatial Correlation Function Simulator

As a part of utilizing a completely simulated data base, access to one with a Markov-structured spatial correlation is required. In order to obtain such a set a white noise process can be transmitted through an appropriate filter.

Let $f(x,y)$, $g(x,y)$ and $h(x,y)$ be the input, the output and the desired filter, then

$$S_{gg}(u,v) = S_{ff}(u,v) |H(u,v)|^2 \quad (B-1)$$

Since $S_{ff}(u,v) = 1$ for white noise and the desired spectral density function, $S_{gg}(u,v) = \frac{2a}{a^2 + u^2} \frac{2b}{b^2 + v^2}$, therefore

$$H(u,v)H^*(u,v) = \frac{2\sqrt{ab}}{(a+ju)(b+ju)} \frac{2\sqrt{ab}}{(b-jv)(b+jv)} \quad (B-2)$$

It then follows that the desired PSF is a one sided exponential i.e.

$$h(x,y) = c e^{-ax} e^{-by} \quad (B-3)$$

let $r_x = \frac{1}{a}$ and $r_y = \frac{1}{b}$ be the filters characteristic length along x and y directions. Then

ORIGINAL PAGE IS
OF POOR QUALITY

$$h(x,y) = c e^{-\frac{x}{r_x}} e^{-\frac{y}{r_y}} \quad x,y \geq 0 \quad (\text{B-4})$$

where c is a normalizing constant providing unity filter gain. Since this filter operates exclusively on digital data, formulation of the problem will be entirely in the discrete domain. Let the filter's length, in pixels, be N_0 . In order to solve for c , the following should hold

$$\begin{aligned} \sum_{m=0}^{N_0-1} \sum_{n=0}^{N_0-1} h(m,n) &= \sum_m \sum_n c e^{-\frac{m}{r_x}} e^{-\frac{n}{r_y}} = 1 \\ &= \left(\sum_m c_1 e^{-\frac{m}{r_x}} \right) \left(\sum_n c_2 e^{-\frac{n}{r_y}} \right) \quad (\text{B-5}) \end{aligned}$$

By equating the individual terms to 1, unity gain will exist along the individual axis as well.

$$\sum_{m=0}^{N_0-1} c_1 e^{-\frac{m}{r_x}} = c_1 \left(1 + e^{-\frac{1}{r_x}} + e^{-\frac{2}{r_x}} + \dots + e^{-\frac{N_0-1}{r_x}} \right)$$

let $\rho_x = e^{-\frac{1}{r_x}}$, the adjacent sample correlation, then

$$\sum_{m=0}^{N_0-1} c_1 e^{-\frac{m}{r_x}} = c_1 \left(\frac{\rho_x^{N_0} - 1}{\rho_x - 1} \right) = 1$$

therefore

$$c_1 = \frac{\rho_x^{N_0-1}}{\rho_x^{N_0} - 1} \quad (\text{B-6})$$

similarly, defining $\rho_y = e^{-\frac{1}{r_y}}$ as the adjacent line correlation,

$$c_2 = \frac{\rho_y^{-1}}{\rho_{y_0}^{-1}} \quad (B-7)$$

but

$$c = c_1 c_2$$

then

$$c = \left(\frac{\rho_x^{-1}}{\rho_{x_0}^{-1}} \right) \left(\frac{\rho_y^{-1}}{\rho_{y_0}^{-1}} \right) \quad (B-8)$$

Since the exponential filter in addition to generating a pixel-to-pixel correlation alters the statistical properties of the input as well, a knowledge of that effect is necessary. Let the input to this filter consist of N random processes corresponding to N spectral bands. The input and output are related by the following discrete convolution;

$$g(m,n) = \sum_{i=0}^{N_0-1} \sum_{j=0}^{N_0-1} f(i+m,j+n)W(i,j) \quad (B-9a)$$

where

$$W(i,j) = h(N_0-1-i, N_0-1-j) \quad (B-9b)$$

and no subscript on $g(m,n)$ and $f(m,n)$ designates any two corresponding bands. The output spatial correlation is

C-4

then given by

$$\begin{aligned}
 R_{gg}(\tau, \eta) &= E\{g(m+\tau, n+\eta) g(m, n)\} \\
 &= \left\{ \sum_{i=0}^{N_0-1} \sum_{j=0}^{N_0-1} \sum_{k=0}^{N_0-1} \sum_{l=0}^{N_0-1} f(i+m+\tau, j+n+\eta) \times \right. \\
 &\quad \left. f(k+m, l+n) W(i, j) W(k, l) \right\} \quad (B-10)
 \end{aligned}$$

Moving the expectation inside;

$$E\{f(i+m+\tau, j+n+\eta) f(k+m, l+n)\} = R_{ff}(i-k+\tau, j-l+\eta) \quad (B-11)$$

Since $f(x, y)$ is a white noise process:

$$R_{ff}(\tau, \eta) = \begin{cases} 0 & \tau, \eta \neq 0 \\ \sigma_f^2 & \tau, \eta = 0 \end{cases} \quad (B-12)$$

Therefore, $R_{gg}(\tau, \eta)$ is non-zero if the following is satisfied

$$\begin{cases} i-k+\tau = 0 \longrightarrow i = k-\tau \\ j-l+\eta = 0 \longrightarrow j = l-\eta \end{cases} \quad (B-13)$$

substituting (B-13) in (B-10)

$$R_{gg}(\tau, \eta) = \sigma_f^2 \sum_{k=l}^{N_0-1} \sum_{l=\eta}^{N_0-1} W(k, l) W(k-\tau, l-\eta) \quad (B-14)$$

ORIGINAL PAGE IS
OF POOR QUALITY

From (B-9b)

$$\begin{aligned}
 W(k, \ell) &= c e^{\frac{k-(N_o-1)}{r_x}} e^{\frac{-(N_o-1)}{r_y}} \\
 &= W(0,0) \rho_x^{-k} \rho_y^{-\ell}
 \end{aligned} \tag{B-15a}$$

where

$$W(0,0) = c e^{\frac{N_o-1}{r_x}} e^{\frac{N_o-1}{r_y}} \tag{B-15b}$$

$$\begin{aligned}
 R_{gg}(\tau, \eta) &= W^2(0,0) \left[\sum_{k=\tau}^{N_o-1} \rho_x^{-2k+\tau} \right] \left[\sum_{\ell=\eta}^{N_o-1} \rho_y^{-2\ell+\eta} \right] \sigma_f^2 \\
 &= W^2(0,0) \rho_x^{\tau} \rho_y^{\eta} \left(\rho_x^{-2\tau} \frac{\rho_x^{-2(N_o-\tau)} - 1}{\rho_x^{-2} - 1} \right) \left(\rho_y^{-2\eta} \frac{\rho_y^{-2(N_o-\eta)} - 1}{\rho_y^{-2} - 1} \right) \sigma_f^2 \\
 &= W^2(0,0) \rho_x^{-\tau} \rho_y^{-\eta} \left(\frac{\rho_x^{-2(N_o-\tau)} - 1}{\rho_x^{-2} - 1} \right) \left(\frac{\rho_y^{-2(N_o-\eta)} - 1}{\rho_y^{-2} - 1} \right) \sigma_f^2
 \end{aligned} \tag{B-16}$$

The variance of the output process is therefore given by

$$\sigma_g^2 = W^2(0,0) \left(\frac{\rho_x^{-2N_o} - 1}{\rho_x^{-2} - 1} \right) \left(\frac{\rho_y^{-2N_o} - 1}{\rho_y^{-2} - 1} \right) \sigma_f^2 \tag{B-17}$$

This result approaches the continuous version for large N_o , r_x and r_y i.e. $\rho_x^{-2N_o} \gg 1$ and $\rho_x^{-2} - 1 \approx \frac{2}{r_x}$ and similarly for

ρ_y . Under these conditions

$$\sigma_g^2 / \sigma_f^2 \approx 1/4 r_x r_y \quad (\text{B-18})$$

Since the primary purpose of this filter is the generation of a pixel-to-pixel correlation of some prescribed value, the following should hold

$$R_{gg}(1, \eta) = \rho_x \sigma_f^2 \quad (\text{B-19})$$

$$R_{gg}(\tau, 1) = \rho_y \sigma_g^2 \quad (\text{B-20})$$

let $\tau=1$ and $\eta=0$ in (B-16). Using the approximation $\rho_x^{-2(N_0-1)} - 1 \approx \rho_x^{-2(N_0-1)}$ it immediately follows that

$$R_{gg}(1, \eta) \approx \rho_x \sigma_f^2 \quad (\text{B-21})$$

and similarly for $R_{gg}(\tau, 1)$.

The next topic is the across-band statistical and spatial effects the exponential filter might have had on the multispectral white noise. Following an exact analog of the derivation presented so far, the crosscorrelation function for any two bands at the output, g_i and g_j is given by

$$R_{g_i g_j}(\tau, \eta) = W^2(0, 0) \rho_x^{-\tau} \rho_y^{-\eta} \left(\frac{\rho_x^{-2(N_0-\tau)} - 1}{\rho_x^{-2} - 1} \right) \left(\frac{\rho_y^{-2(N_0-\eta)} - 1}{\rho_y^{-2} - 1} \right) r_{ij} \sigma_{f_i} \sigma_{f_j} \quad (\text{B-22})$$

where the input crosscorrelation function is defined

$$R_{f_i f_j}(\tau, \eta) = \begin{cases} 0 & \tau, \eta \neq 0 \\ r_{i,j} \sigma_{f_i} \sigma_{f_j} & \tau, \eta = 0 \end{cases} \quad (\text{B-23})$$

The band-to-band correlation coefficient at the output is given by

$$r_{g_i g_j} = R_{g_i g_j}(0, 0) / \sigma_{g_i} \sigma_{g_j} \quad (\text{B-24})$$

It then immediately follows that

$$r_{g_i g_j} = r_{f_i f_j} \quad (\text{B-25})$$

i.e. the correlation coefficients has undergone no change under this transformation

ORIGINAL PAGE IS
OF POOR QUALITY